ADVANCES IN DIAPRIID (HYMENOPTERA: DIAPRIIDAE) SYSTEMATICS, WITH CONTRIBUTIONS TO CYBERTAXONOMY AND THE ANALYSIS OF RRNA SEQUENCE DATA

A Dissertation

by

MATTHEW JON YODER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2007

Major Subject: Entomology

ADVANCES IN DIAPRIID (HYMENOPTERA: DIAPRIIDAE) SYSTEMATICS, WITH CONTRIBUTIONS TO CYBERTAXONOMY AND THE ANALYSIS OF RRNA SEQUENCE DATA

A Dissertation

by

MATTHEW JON YODER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Robert Wharton
Committee Members,	Anthony Cognato
	Rodney Honeycutt
	James B. Woolley
Head of Department,	Kevin Heinz

May 2007

Major Subject: Entomology

ABSTRACT

Advances in Diapriid (Hymenoptera: Diapriidae) Systematics, with Contributions to Cybertaxonomy and the Analysis of rRNA Sequence Data. (May 2007)
Matthew Jon Yoder, B.S., University of Waterloo; M.S. Texas A&M University
Chair of Advisory Committee: Dr. Robert Wharton

Diapriids (Hymenoptera: Diapriidae) are small parasitic wasps. Though found throughout the world they are relatively unknown. A framework for advancing diapriid systematics is developed by introducing a new web-based application/database capable of storing a broad range of systematic data, and the first molecular phylogeny specifically focused at examining intrafamilial relationships. In addition to these efforts, a description of a new taxon is provided. Several advantages of digital description, including linking descriptions to an ontology of morphological terms, are highlighted. The functionality of the database is further illustrated in the production of a catalog of diapriid host associations. The hosts database currently holds over 450 association records, for over 500 named taxa (parasitoids and hosts), and over 180 references. Diapriids are found to be primarily endoparasitoids of Diptera emerging from the host pupa. Phylogenetic inference for a molecular dataset of 28S and 18S rRNA sequence data, derived from a diverse selection of diapriids, is accomplished with a new suite of tools developed for handling complex rRNA datasets. Several parsimony-based methodologies, including an alignment-free method of analyzing multiple sequences, are reviewed and applied using the new software tools. Diapriid phylogenetic relationships are shown to be broadly congruent with existing morphology-based classifications. Methods for analyzing typically excluded sequence data are shown to recover phylogenetic signal that would otherwise be lost and the alignment-free method performed remarkably well in this regard. Empirically, phylogenetic approaches that incorporate structural data were not notably different than those that did not.

ACKNOWLEDGEMENTS

Thanks to my family, Eric, Joan and Tim for providing an environment that fostered curiosity (and good food). Many people provided insights, feedback and support and this work would not be possible without them: Robert Wharton, Jim Woolley, Anthony Cognato, Rodney Honeycutt, Joe Gillespie, Andy Deans, James Munro, Lubomir Masner, Johan Liljeblad, Debra Murray, Katja Seltmann, Wylie Harris, Ozlem Altiok, Istvan Miko, Jacques Dubois, Viktor Kolyada, Hans Clebsch, James Dennis, Jonathan Cammack, Karl Roeder, Bill Wells, Tiffani Williams, and members of the Entomology and Ento-CS discussion groups, Wharton and Heraty Labs. The Department of Entomology staff, in particular Robert Jensen and Teresa Gold made my experience as an international student much easier, thank you. Special thanks to Amy Bader. Finally, Abdy and Jbo provided much support and encouragement, virtual and otherwise.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Diapriidae – Introduction and Placement Within the Apocrita

Diapriids are tiny parasitic wasps. They are an abundant, diverse, and cosmopolitan group (Masner, 1993). Most remain unnoticed due to their small size, which averages ca. 2-3mm in length. Well known species are primarily solitary or gregarious endoparasitoids of dipteran larvae and pupae (Chambers, 1971; Hoffmeister, 1989; Masner, 1993). Some species are also known to attack ants (Lachaud and Passera, 1982; Loiácono, 1987) and beetles (Masner, 1993; Brown and Arrington 1967). Immature stages are completely undescribed for nearly all species (but see Silvestri, 1913; Kazimirova and Vallo, 1999; Coon, 2000).

Within the Hymenoptera the Diapriidae are placed in the 'Parasitica' grade of families within Aprocrita. Members of this group are almost exclusively parasitic on insects and other arthropod hosts. Most authors have further placed the Diapriidae in the Proctotrupoidea s.l., a superfamily that is now considered to be paraphyletic (see phylogenies in Dowton and Austin, 2001). Whereas many relationships within the Parasitica are still poorly understood (Whitfield, 1992; Dowton et al., 1997; Gibson, 1999; Ronquist et al., 1999; Dowton and Austin, 2001, Sharkey and Roy, 2002) it now appears that the Diapriidae are perhaps sister (Naumann and Masner, 1985; Dowton and Austin, 2001) to the rare, relictual, families Maamingidae (Early et al., 2001) and Monomachidae (Musetti and Johnson, 2004).

Intrafamilial classification has to-date been based on hypotheses of morphological synapomorphy. Four subfamilies are presently defined, the Ismarinae, Ambositrinae, Belytinae, and Diapriinae. Minimal tribal classification has been proposed for the Ambositrinae (Naumann, 1982), Belytinae (Macek, 1989b) and Diapriinae (summarized in Masner and García, 2002). The Ismarinae are monotypic, with one widespread genus

---

This dissertation follows the format and style of *Systematic Biology*.

Ismarus Haliday. A second currently available name, *Szelenyioprioides* Szabó, is considered by Megyaszai and Thuróczy (1998) to not belong in the Ismarinae. Relative numbers of genera and species by geographic region are listed in Table 1.1.

Literature facilitating the identification of diapriids is biased strongly towards European species (Table 1.1). Because of this geographic bias, the estimated large number of undescribed species and the relative paucity of modern keys to species, diapriid taxonomy frequently operates at the superspecifc level (i.e. specimens are commonly not identified beyond genus). Older syntheses (Ashmead, 1893; Kieffer, 1910, 1916) are now considered to be of historical value (see comments in Macek, 1989b; Masner, 1993; Masner and García, 2002), though some (e.g. Foerster, 1856) still provide valuable insights. Several more recent works have provided generic and species treatments in the form of annotated regional keys. Of particular use in this regard are Nixon (1957, 1980), Macek (1989b), Naumann (1982) and Masner and García (2002). Identification (and taxonomic problems) in general are further confounded by the strong sexual dimorphism exhibited by nearly all diapriid species. This frequently necessitates separate keys for males and females, if the different sexes can be associated at all. A final roadblock to taxon identification and delimitation is the marked reduction or loss of morphological features exhibited by many diapriids. As a result of this loss, the species tend to be uniform in appearance within a genus and hence difficult to differentiate. This problem is not unique to diapriids but is commonly seen in other clades of Parasitica (e.g. Platygastroidea, Chalcidoidea) (Gibson, 1985; Dowton et al., 1997; Campbell et al. 2000).

TABLE 1.1. Taxonomic summary of diapriid diversity by biogeographical region. Format for regions: [estimated total genera in region]/[estimated total species names described from region]. Totals are in the format [total described genera/total described species], fossils (Fos) are not included in regional counts. Totals represent currently accepted valid names only. Data are taken from Johnson (1992, 2003), with >120 additional data-points from literature published post 1992. AUS – Australian; ETH – Ethiopian; NEA – Nearctic; NEO – Neotropical; ORI – Oriental; PAL – Palearctic; Fos. – Fossil; * - Species originally described from multiple regions are not included in regional totals.

	AUS	**ETH**	**NEA**	**NEO**	**ORI**	**PAL**	**Fos**	**Totals**
Ambositrinae	15/88	-	1/1	4/8	-	-	1/1	21/98
Belytinae	8/15	7/20	15/160	22/29	9/40	32/415	9/13	58/692
Diapriinae	24/125	26/118	23/189	44/218	26/138	34/414	5/7	116/1209(+10)*
Ismarinae	1/1	-	1/6	1/12	-	1/7	-	1/26(+3)*
totals	48/229	33/138	40/356	71/267	35/178	67/836	15/21	2038

Increased accessibility provided by modern keys should encourage scientists to use diapriids as model organisms for a wide range scientific studies. Biologically, there are a number of evolutionary transitions that may be of interest including those from solitary endoparasitism to gregarious parasitism, and transitions to myrmecophily (Huggert and Masner, 1983), termitophily (Naumann and Masner, 1980), or semi-aquatic biologies (Masner and García, 2002). Also of interest is the potential co-evolution between diapriids and Diptera, as there appear to be at least some preliminary correlation between purportedly basal diapriids and the lower Diptera (Belytinae and Nematocera) and derived members with higher Diptera (*Trichopria* and Cyclorrhapha) (see Hoffmeister, 1989). Parallel evolution between diapriids and ants may also have occurred; however, these relationships are even less clearly understood (Masner and García, 2002). Careful study of these relationships will require a robust phylogeny. Distribution patterns of diapriid species, genera, tribes and subfamilies are also of potential interest to systematists. The Ambositrinae, for example, is primarily a southern Hemisphere, old world subfamily (Masner, 1961, 1969; Naumann, 1982) exhibiting a classic Gondwanan distribution. Biogeographical data will ultimately aid in the construction of a robust hypothesis for the early evolution of the family. Broad-scale biogeographical patterns are further of interest to those studying the evolution of communities or faunal distributions which exhibit similar patterns.

Historically, the Parasitica have been of particular interest because many species are useful in the field of biological control. This has not been the case for diapriids, which for various reasons are generally not considered to be potential biological control agents. One such reason is clearly taxonomic impediment, i.e. the absence of expertise (generally due to lack of study) available to identify organisms to a scientifically meaningful level. For example, species of *Trichopria* are known to attack a large number of pestiferous species, but in the literature (>20 references) are very frequently identified either as *Trichopria* sp. or undescribed. Greathead (1986) states that no diapriid was used/introduced for biological control up until 1986. He overlooked at least one case, the introduction to Hawaii of *Coptera silvestrii* by Silvestri (1913). Since then at least two other species have been introduced for control purposes (Vallo, 1989; Hellqvist, 1994),

and several others have been studied directly or indirectly for their potential use (Legner et al. 1967; Legner and Olton, 1968; Roberston, 1987; Kazimorova and Vallo, 1992; Sivinski et al., 1998). Other studies have noted that some diapriid species attack useful biological control agents, and thus potentially hinder biological control efforts (Coon, 2000). Regardless of their use as beneficial organisms in biological control, it is clear that they have a potential impact on control programs, and as such their study is important.

The role of diapriids in community ecology is also poorly understood, even though diapriids are ubiquitous to a huge variety of environments. Perhaps the largest studies on their ecological importance are those of Garbarczyk (1981) and Ulrich (2000), both of which found correlation between abiotic and biotic factors and various measures of diapriid diversity and presence. Diapriids are commonly recovered in studies of arthropods associated with commercial animal production, presumably as parasitoids of Diptera in manure. They are also observed in richly organic microhabitats, or associated with fungi (Huggert, 1979). However, levels of recovery are frequently very low (e.g. Smith et al., 1987; Hoggsette et al., 1994). Their niche in natural ecosystems is hard to deduce based on these limited data.

Though major gaps in our understanding of diapriid taxonomy exist, particularly in the subfamily Belytinae (but see Wall, 1967; Hellèn, 1964; Macek, 1989ab, 1993ab, 1994, 1995abcd, 1996, 1997abc, 1998), the field is relatively accessible to those seeking to make immediate contributions. Historically this is largely due to comprehensive nature of Kieffer's earlier works (see discussion in Naumann, 1982; Macek, 1989ab; Masner, 1993), the complete catalogs of all diapriid taxonomic literature (digital and paper) of Johnson (1992, 2003), and the more modern larger scale efforts of Naumann (1982, 1987, 1988), Masner and García (2002), Masner (1976), Notton (2004), and Nixon (1957, 1980).

Dissertation Overview

Given the relative accessibility of the group this dissertation is largely an attempt to begin to meld technologies available to modern systematic biologists with diapriid taxonomy and phylogeny. Many of these technologies broadly full under the relatively

recently defined field of "Biodiversity Informatics". In this regard the work presented here should be seen as the first step of a long term approach to advancing diapriid systematics. To this end the dissertation is presented as follows.

Chapter II introduces "mx", a web-based content management application for systematic biologists. The mx project was initiated to provide a web-based platform for storing and disseminating systematic-related data in general and more specifically for the diapriid research in this dissertation. The database acts as a repository for nearly all the data underlying the research presented here. The database's general usefulness has lead a number of additional researchers to employ its functionality for their research, and the project is being aggressively promoted as an open-source solution with the hope that a community-based adoption will occur.

Chapter III provides a simple example of using mx to create a taxon page, and illustrates the database's capability to link descriptive text to an ontology of morphological terms. It represents the first published taxonomic description using the mx system.

Chapter IV describes and provides an example of using mx to store biological relationships among taxa, in this case the host-parasite relationships for the family Diapriidae.

Chapter V presents the first molecular phylogeny for the Diapriidae. Using molecular data and a suite of parsimony-based tools both a phylogeny and exploration of the use of rRNA molecules in systematic research is provided.

Chapter VI concludes the dissertation with some reflections on the process of modern systematics/taxonomy given the effort undertaken in this dissertation.

CHAPTER II

MX – A COLLABORATIVE CONTENT MANAGEMENT SYSTEM FOR SYSTEMATIC BIOLOGIST

Overview

The lack of available collaborative tools for use by systematic biologists is arguably a hindrance to large scale initiatives like NSF's Assembling the Tree of Life and Planetary Biodiversity Inventories initiatives. We present the application 'mx' (short for 'matrix'), a completely web-based content management system for use in systematic biology. This free, open-source software is built using a MySQL database and the Ruby on Rails web-application framework. The project contains over 50 related tables tracking a wide variety of systematics-related data including taxon names, specimens, collecting events, biological associations, primary literature, morphological and molecular character data, and images. A single installation of mx can support multiple projects, each with multiple users, allowing for long distance collaborations (e.g., in matrix scoring). The system uses as its core data object an OTU (Operational Taxonomic Unit), which allows for taxonomic concepts to be temporarily independent of formally governed names while a revision or biodiversity study is undertaken. An installation of mx is currently being used for a range of projects and tasks, including matrix management, taxonomic catalogs, biological relationship tracking, and taxonomic revisions. The projects are reviewed herein.

Introduction

The importance of the earth's biodiversity has recently been underscored by a range of major funding opportunities (e.g., the National Science Foundation's Partnership for Enhancing Expertise in Taxonomy (PEET), Planetary Biodiversity Inventory (PBI), and Assembling the Tree of Life programs) and large scale initiatives such as the Global Biodiversity Information Facility (GBIF, http://www.gbif.org/), Integrated Taxonomic Information System (ITIS; http://www.itis.usda.gov/), and Species 2000 (http://www.sp2000.org/). These programs are all at least in part a response to what has been termed the "taxonomic impediment" (Taylor 1983; Rodman and Cody 2003). This impediment is largely a reflection of the large numbers of earth's species that remain undescribed (and the urgency that potential extinctions place on this work), coupled with the relative paucity of experts and resources available to accomplish this task. The enormity of the former task has led for calls to re-assess the methods by which taxonomy has been accomplished to date, including a call for and implementation of a wide range of software-based tools, such as electronic taxonomic catalogs, image databases, taxon Web pages, and DNA barcoding initiatives.

The evolution of these tools follows a number of pathways. On one end of the spectrum are custom-built databases used daily for the immediate needs of the practicing systematist (e.g., collecting event labels captured in a text file), and on the other end are global collaborations with massive infrastructures (e.g., GBIF; and Ecoport at http://ecoport.org/). The latter are based on or provide schemas (e.g., the Darwin Core) that represent standards for data to be maximally useful, manageable and exchangeable. The former are strictly utilitarian and may or may not even include a schema. The needs of the *practicing* systematist, i.e. one who is actively describing new taxa or hypothesizing new evolutionary relationships, are likely not fully met by a strategy from either extreme. Complex schemas do little to increase the efficiency of a taxonomist unless practically implemented, while simple text files are less efficiently queried, proofed and collaborated on. The opposing ends of this spectrum are drawing together due to recent software advances that allow for the relatively simple implementation of

complex database schemas and the existence of widespread standards such as those overseen by the Taxonomic Databases Working Group (TDWG, http://www.tdwg.org).

We describe here the 'mx' content management system (CMS), a project that seeks to be intermediate in scope, between simple lists and complex implementations. This CMS provides a large underlying database schema housed in a relatively simple front-end application. Broadly stated, its goal is to store and manage the information generated during the process (sensu Franz 2005) of modern revision or monography. We believe this process necessarily includes both traditional taxonomic efforts (e.g. alpha-taxonomy) and phylogenetic inference. The database structure of mx is sufficiently parsed such that data in it may ultimately be translated to larger archives (e.g., Tree of Life Project (Maddison and Schulz 2007) and Morphbank (http://morphbank.net). We seek to provide in mx a balance between model or data-structure (e.g. Dallwitz and Paine 2004; Pullan et al., 2005) and existing application.

The evolution of mx is following several broad principles that define the project and shape its development:

- the desire to ultimately capture all data pertinent to a modern taxonomic revision during the revisionary process such that simultaneous publication of all data in print and electronic forms is possible
- the development of a web-based multi-user front end whose components may also be used (or translated relatively seamlessly) for public display of data
- the implementation of a taxon-concept based OTU ('O'perational 'T'axonomic 'U'nit) as the core data object, such that published and unpublished (working) concepts can be managed
- the use of freely available software (including the mx sourcecode) in an aggressively open-source framework

While these characteristics are all variously found in other applications, and there is a long history of informal discussion on these topics (see the TDWG website at http://www.tdwg.org), there are remarkably few, if any, available applications (including

mx) that fully integrate *all* of the principles listed above. A recent web-based survey instantiated on the Taxacom listserv was aimed at discovering those efforts that considered themselves web-based content management systems; it resulted in 14 responses (http://vsmith.info/node/17987). While accession management is perhaps a subset of a fully featured content management system there are many overlaps among the two, and a survey of collection management software by Berendsohn et al. (2003) returned 24 responses. While the administrators of these surveys acknowledge the informal nature of their data gathering, it seems clear that the diversity of biodiversity content management systems is relatively low.

One problem, which the NSF PEET grants seek to address, is the loss of taxonomic knowledge when an expert retires or is lost (Rodman and Cody, 2003). PEET projects are generally designed to recover this knowledge by apprenticing young workers with seasoned taxonomists. Taxonomic experts inevitably publish much less than they know (not a fault, but rather a limitation of time). What is not published may be passed on to others, but this does not always happen: in essence the human database can be lost. If taxonomists capture thoughts pertaining to a majority of the taxonomic concepts they recognize as they recognize them, then much of this information can be of use for future generations. To date this has usually been done in lab and field notebooks, if at all. This problem of "losing" the concept in question is exacerbated by the trend in systematics to emphasize phylogeny over classification (see Franz, 2005). Mx implements a simple way to make this data capture possible in an electronic framework, providing easily defined "concept" and "content" categories. Tracking taxon-concepts is particularly important in biodiversity studies, which may involve any number of taxonomists or parataxonomists who spend a large amount of time identifying morphospecies but little time recording how those morphospecies were delimited. In this sense an OTU in mx acts in part as the "potential taxon" of Berendson (1995) or the "taxonomic concept" of Franz (2005). It is also important to capture data as "potential taxa" for those lifetime experts whose primary work is the study of a clade of organisms. For example, taxonomists working on speciose taxa (e.g. Hymenoptera) may recognize

many more species than they formally describe during their lifetime. If detailed notes pertaining to these taxa are tied to identifiers and included in an electronic database then subsequent generations of workers gain testable hypotheses based on the accumulated experience of past workers rather than drawers of cryptically arranged and labeled specimens. Tracking taxonomic concepts is not novel, and its underpinnings lie in philosophical debates concerning the delimitation of species concepts. Recent papers (e.g. Franz, 2005, Kennedy et al. 2005) address the need for a concept based database structure, wherein a framework is developed to capture and exchange the complexity associated with a taxonomic concept. Mx provides a mechanism to achieve this goal but focuses on assisting with taxonomic and phylogenetic progress rather than capturing historical concepts in the comprehensive framework described in Kennedy et al. (2005).

Solving the taxonomic impediment requires maximum efficiency of the taxonomic process (Erwin, 2000). The processes that lead to electronic and paper publication of taxonomic descriptions should be identical - i.e., a single process which allows for the processing of results prior to paper and electronic publication is more efficient than a two part process (i.e., finishing a paper revision then translating to electronic form). Mx is designed, in part, to do exactly that. A paper publication should be a trivial bi-product of an "electronic revision". Efficiency can also be maximized within the application itself. We seek to do this in mx by using a web based interface so that components designed for use in day-to-day data capture and manipulation are readily available for use in (or simple translation to) the publicly accessible side of the application.

Several other initiatives provide applications or goals similar to those of mx (e.g. BioCorder, http://www.biocorder.org/; EDIT, http://www.e-taxonomy.eu/), but the number of initiated projects that are currently operational, open-source, and that encourage modifications by someone other than the initial developers of said applications are a minority (albeit growing). This is likely due to 1) the desire to retain central authorship and thus recognition (citation) for the project, 2) the desire to receive additional funding or grants for development of the project, 3) the thought that scientific

applications require more control and oversight because of the rigor inherent to the scientific method, and 4) the relative uncertainty associated with an open-source development process. While these are legitimate and important concerns, they are not necessarily reasons to withhold source code. One advantage to open source projects is that users can share incremental improvements. If a user creates a new feature (e.g., new search functionality) it becomes available to all the other users of mx. Because of this potential we elect to present the project and make available its code base early in its development, with the hope that the input of others will ultimately provide a more rapidly evolving tool with expanded capabilities for the systematics community. In this light we do not proclaim to be a final solution but rather a working test bed for ideas proposed in past works (particularly Franz 2005).

Materials and Methods

Target Users and Usage

While mx is flexible enough to handle a wide range of data, its interface and data-structure are primarily developed for use by revisionary systematists or small teams of systematists who are faced with treating a large number (> 30) of undescribed taxa. An installation of the database can also support labs with multiple users, each working on a smaller project (e.g., managing matrices, tracking voucher specimens, creating taxonomic catalogs). Examples of current usage are listed in Table 2.1. Readers are encouraged to visit the homepage(*http://hymenoptera.tamu.edu/wiki*) to visualize the application and track ongoing work. .

TABLE 2.1. Sampling of projects currently maintained in a single instance of mx, a content management system for biological systematists.

Project	Overview
Diapriid Systematics	MJY and several collaborators are using mx to store observations, images, keys and notes pertaining to long-term revisionary projects aimed at addressing the taxonomic impediment as it pertains to the Diapriidae (see http://www.diapriid.org)
Diapriid Hosts	This project seeks to capture all know host information for the Diapriidae (Hymenoptera) and represents an international collaboration of 4 co-authors (see http://www.diapriid.org).
Evaniid Systematics	mx acts as the repository for Catalogous Evaniidorum (http://evaniid.tamu.edu) the digital version of Deans (2005), a catalog of evaniid names and literature. Mx is also used to track vouchers for ongoing molecular analyses (phylogenetics) and biodiversity studies.
Phylogenetics of Xyloborini	The HISL lab (~5 members, http://hisl.ent.msu.edu/) is using mx to catalog its material, collecting event information and sequence related data. It has also transcribed several large taxonomic catalogs and the references and keywords therein to mx.
American Entomological Institute Type Catalog	A complete catalog of the Townes ichneumonid (Hymenoptera: Ichneumonoidea) types housed at the American Entomological Institute (http://www.amentinst.org/).
Ichneumonoid Systematics	Morphological characters and matrices for a phylogenetic revision are being managed for several masters-level theses and a larger species-level revision.
Hymenopteran Molecular Systematics	mx houses several projects on hymenopteran molecular systematics. For example vouchers and extracts for the Hymatol PCC project (http://ceb.csit.fsu.edu/ronquistlab/PCCP/) project and primer sequences for rRNA secondary structure projects.
Hymenoptera Morphological Ontology	Members of the Hymenoptera Tree of Life (http://www.hymatol.org/) team are editing a list of over 1200 terms (and growing) and definitions pertaining to Hymenoptera morphology in an effort to unify and update usage.
Chalcidoid Morphological Supermatrix	The Heraty lab at the University of California Riverside, is leading a team of researchers in this review, annotation and study of over 400 morphological characters (> 800 states) for parasitic wasps in the superfamily Chalcidoidea.
A multiple-entry key to Families of Apoidea	A matrix of over 50 morphological characters underlies this key to the 12 bee families.

Application

There are two major components to mx: the underlying MySQL (http://www.mysql.com) database and the Ruby on Rails (http://www.rubyonrails.org/) application. The former is stored as a set of SQL statements so that the database can be (re)built by any user with a MySQL database server. Usage of another database engine (e.g. Postgres, DB2) should be trivial, with only minimal translations needed. Ruby on Rails is a relatively new Web-application framework and is somewhat unique in that it enables extremely rapid development, yet results in concise, intelligible code that is easy to maintain and extend. Rails also provides the developer simple interfaces to AJAX and other dynamic Web building tools, easing the process of creating a responsive Web interface. As Rails is object-relational it meets many of the "engineering considerations" of Morris et al. (2002). The project is under version control using SVN (http://subversion.tigris.org/), and sourcecode is available at Sourceforge (https://sourceforge.net/projects/mx-database/). Image manipulation is handled by the ImageMagick (http://www.imagemagick.org) toolkit. The project's homepage and source code can be found through links at http://hymenoptera.tamu.edu/wiki.

The database structure (simplified in Fig. 2.1) is influenced by a wide range of databases and utilities, various incarnations of which have been in use over the course of the past 7 years. Though highly modified (primarily simplification) from the original, the basis for the character and matrix table-structure is DeltaAccess (Hagedorn, 2003), a RDBMS built in Microsoft Access ™ that integrated much of the functionality defined in the Delta specification (Dallwitz and Paine 2004). The present table-structure, however, is greatly simplified from the DeltaAccess implementation. Much of the remaining underlying table structure is derived from the Texas A&M Insect Collection Database (Yoder, Oswald unpublished). Aspects of the Metacanthomorpha database (Dettai et al. 2004) have also been included in the schema but remain undeveloped. The mx database structure contains all of the DarwinCore 2 (http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome/) fields for specimens, and functionality to export data in conjunction with the DarwinCore 2 namespace to Berkeleymapper is already available. The majority of the schema was in development prior to Kennedy et al (2005), however many of the features they discuss, in terms of concept based management, are available in practice or concept in mx. For instance, specimens may be given any number of determinations and/or unique identifiers, allowing for the containment in any number of concepts ('OTUs' in mx).

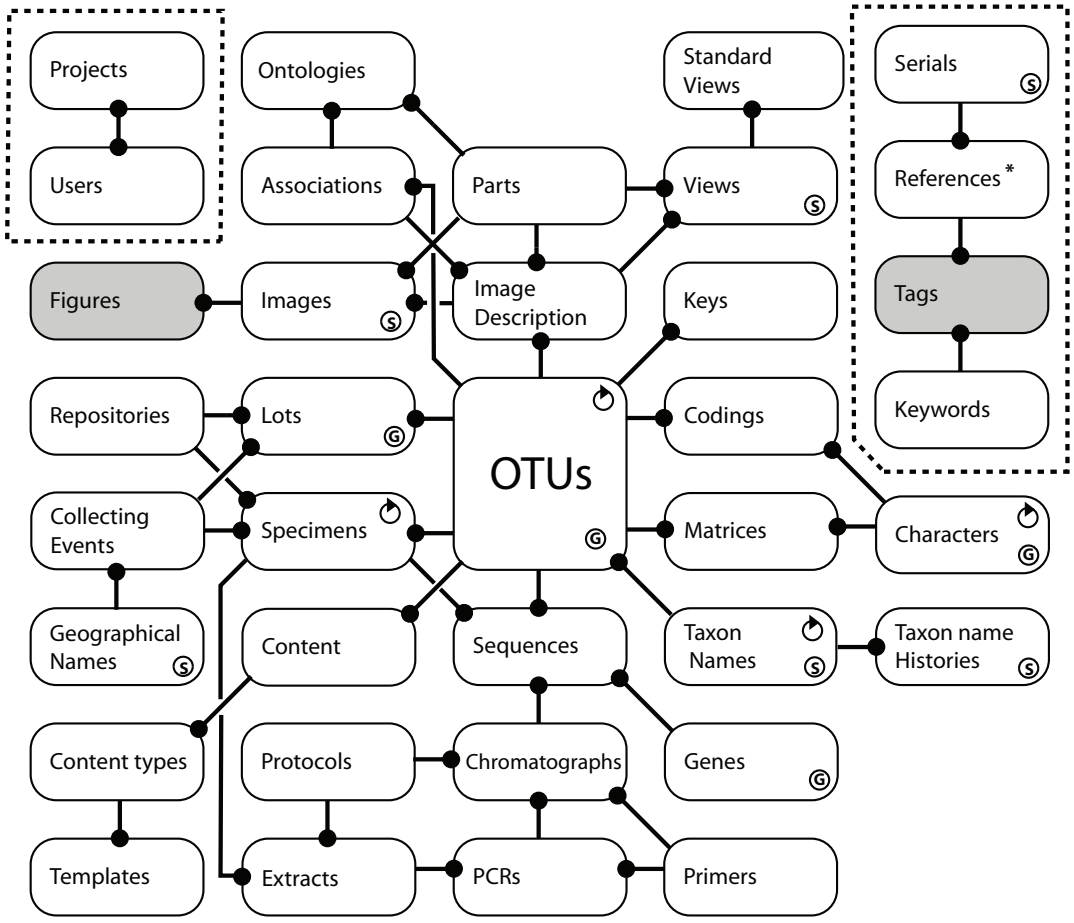

FIGURE 2.1. Generalized overview of the 'mx' content management system for systematists. Grayed objects are universally applicable (potentially linked) to any other object. Relationships with the "many" side indicated with a shaded ball. Objects with a circled "G" may be grouped into named groups which can be used in management and display functionality. A circled "S" indicated objects that are shared across projects. * - References are linked to a majority of other objects. Self-referential (parent-child) relationships are indicated by a circular arrow.

Key Features for Users

Currently implemented features are discussed below, though some are relatively rudimentary. Numerous additional features are omitted from this list (for brevity) and are visible by examining the application or source code.

The interface to mx is web based, which allows for notes and data to be taken anywhere a browser is accessible. This is particularly useful for remote data capture and retrieval, for instance while visiting museum collections. Similar to applications like MorphoBank (O'Leary and Kaufman, 2007) and Morphbank (http://morphbank.net), a single installation of mx can support teams of researchers from different institutions, facilitating remote collaborations. For example, two or more people can work simultaneously on the same revision or score different parts of a morphological matrix from different computers. Furthermore, any number of individual projects can be created within a single installation of mx; this allows a single developer or database expert to efficiently support a number of labs. It is also possible to use a local (non-web based) instance of mx, though this functionality is not currently optimized.

The core OTU object provides the means to capture data during a project in process without having to formally identify the taxon in question. The OTU class essentially allows for a level of uncertainty to be identified and incorporated into the revisionary process. This is particularly critical to revisions or biodiversity studies that may handle formally unidentified taxa for long periods of time prior to publication. This is also important when taxonomic concepts published in the past are in fact misinterpretations, and it is necessary to identify these as such. In mx these past interpretations (or hypotheses) can be included as full OTUs in their own right and then synonymized with current concepts.

Usage of the term "OTU" has historically been linked to phenetic (Sokal and Sneath, 1973) schools of thought. While we do not encourage a phenetic approach to taxonomy, we retain the usage of OTU in mx to emphasize "operational" - i.e., an OTU is a circumscription of ideas with which something meaningful can be done (e.g.,

generated hypotheses). OTUs in mx are hypotheses of taxonomic entities. These hypotheses must be supported with observations taken on physical specimens. In mx, data supporting the hypothesis represented by a particular OTU are captured primarily as direct links to specimen records, descriptive statements placed in any number of categories ("content types" in mx), images, keywords, or as matrix character codings.

Specimens in mx are individual organisms or parts thereof, with parts being attached to bodies in parent-child relationships. Specimens can be given any number of determinations (ties to OTUs) and or unique identifiers. They may be associated with a collecting event, which is further linked to geographic authority tables. Lots (records identifying multiple individuals of the same OTU) can also be tracked in mx.

Mx employs a text-content system very similar to that used in the Tree of Life project (Maddison and Schulz, 2007), wherein any number of content types (text fields) can be defined, and grouped in any number and way in 'templates'. Images may be attached to each unique OTU/content type combination, and uniquely annotated. Templates for taxon pages, rough drafts, or other informative pages are easily composed in this system. Example pages are available at http://www.diapriid.org and http://evaniid.tamu.edu.

All core objects (records with unique IDs) in mx can be enhanced by attaching "tags" or "figures" to them (Fig. 2.1). Tags are keywords that are attached to content. Tags can be used in conjunction with a reference, for instance a "keys" keyword could be used to link taxon names to references that contain keys for each taxon. Any number of keywords can be created and used as tags. Figures are similar to tags but they relate images to core objects rather than keywords. The images used in figures (and elsewhere) can be managed directly in mx as locally uploaded files or as links to images and thumbnails managed by Morphbank (http://morphbank.net). Morphbank is a major NSF funded effort to archive and provide tools for working with biological images of all types. Morphbank images are easily integrated by simply creating a new 'Morphbank' image in mx, which is simply a pointer to an image managed in Morphbank.

A simple ontology editor is also available in mx. The ontology table relates terms via any number of user definable categories (e.g. "is a", "synonym of"). Parts (like

all database objects) may be figured and referenced. It is trivial to link any block of text in mx (implementations exist for taxon descriptions, character descriptions and key couplets) to the ontology, greatly enhancing the informativeness of these data (Yoder, 2007).

For morphological matrices mx uses a relational model (see Nixon et al. 2001) to store character and matrix data. Matrices in mx are defined as a combination of OTUs and characters and are strictly utilitarian. This means that once an OTU is coded for a character it is coded for that character in all matrices in which that combination of character and OTU are included. Multiple matrices can be created and managed by using a flexible system of adding OTUs and characters in groups or individually, and deletion or modification of character or OTU will not alter the codings of OTUs therein. The management of larger matrices is easily accomplished through the use of smaller matrices. For instance, separate matrices could be coded by separate users for logically different morphological groups (e.g. head, arms, legs) and then combined into one larger matrix by including all character groups. Characters may be grouped into any number of arbitrary groups. This allows for aggregation of characters into logical (e.g. "head" or "thorax") or subjective ones ("informative" or "uninformative") classes. All characters may be tied to references, and a character can be "synonymized" by pointing it at another character. Numerous other functions are available including one-click coding of OTUs for a given matrix, merging of character states, and trivial exchange or update of character state labels. Matrices are exportable to TNT, WinClada and Nexus formats.

Morphological matrices can be used as multi-entry keys in mx. The multi-entry key engine provides a number of different ways to view and operate the key and can move flexibly back and forth through the decision making process. Traditional bifurcating keys may also be created, illustrated and linked to OTUs (see Yoder, 2007).

When one decides to formally describe or identify an OTU it can be linked to a taxonomic name - i.e., one that is consistent with the ICZN (International Commission on Zoological Nomenclature, 1999). A single tree of taxonomic names is shared across all projects. Administrative level options allow for per project control of both the

visibility and modification of taxon names. For instance a user may be able to see all members of a family, but only edit or add members of a given tribe within that family. Taxon names can then be exported in an ITIS identity file format.

While the primary focus has been on developing the features listed above, mx also contains a range of other features important to systematists. A DNA sequence database is included in mx. A range of sequencing related data (vouchers, extracts, PCRs, sequences, primers, gene names, protocols, and gene groups) can be stored in mx, and unaligned sequences can be finely managed (grouped by OTU group or gene group) and exported in a variety of unaligned formats (e.g., FASTA). Nearly all objects in mx can be tied to references. An "associations table" allows for the recording of biological relationships among OTUs.

Finally, a series of functional features have been developed to aid users of mx. Mx generates help links to a centralized wiki such that help documents can be easily created, updated, and maintained. It also contains a simple in-application set of tables for storing help documents for each central object type. This allows help to be written and tied to the category of interest as questions arise. A 'news' feature allows mx administrators to communicate with users by posting notices that expire automatically after a set period of time, and allows users within a given project to communicate within their project, or when data are made public externally. Mx also allows for a number of user customizations, allowing certain content to be temporarily hidden or shown.

Key Features for Developers

The rapid development of the mx project and handling of a relatively large table-structure was made possible in part by the Rails application framework (http://www.rubyonrails.org/). Indeed, the usability of Rails makes it (and by association mx) a useful tool for courses that introduce bioinformatics databases to systematists in training. Rails integrates several interactive client-server technologies relatively seamlessly, so that components like popup style forms, auto-completing text boxes, and real-time searches are readily incorporated. New tables (objects in Rails) can be rapidly

implemented in the application using Rails' scaffolding feature, which automatically generates a basic user interface. Source code for mx is under version control so developers can check out and independently modify or add features locally, then post them back to be incorporated into the trunk (source code base). The ability to share features and bugfixes should lead to more rapid development of the core application.

Present State of Development

As mx is rapidly being developed interested parties are encouraged to visit the home-page wiki (http://hymenoptera.tamu.edu/wiki) and Sourceforge repository (http://sourceforge.net/projects/mx-database) to see the most recent status of the project. A range of screen captures and news can be found therein, along with information on installing, developing, and using mx. Guides have been written for the matrix-related functionality of mx and these are also available on the wiki.

The current database schema contains the large majority of tables planned for the project, and user interfaces are available for nearly all of them. A production installation of mx has been in use since April 2005, with over 80 registered users variably involved in 40 projects. Mx is evolving with increased use, and new features are being developed as they are needed. This interaction between users and developers has driven the development of the project towards immediately useful features. Planned major additional features include: 1) rapid specimen handling (i.e., with label barcoding in mind- the ability to quickly and efficiently add, search, and update multiple specimen records with a single operation), 2) interaction with Morphbank such that mx-managed images and their associated data can be archived easily, 3) the integration with GBIF and other federated organizations using pyWrapper/TAPIR (http://www.pywrapper.org/), and 4) the development of a work-flow module that chains day-to-day tasks in customizable combinations.

Discussion

Until recently, efforts to revise and develop classifications of groups of organisms have largely been accomplished by taxonomists working alone. As evidenced by the major funding initiatives mentioned t, the current trend is changing towards team-based efforts. While these types of collaborations are possible via piecemeal approaches it is strongly desirable to capture and work on data through a centralized application (e.g., web-based solutions). Taxonomic research has a large cataloging component to it, with revisions and monographs essentially representing compendiums of our current knowledge of a group of organisms. A huge number of existing CMSs are available to those seeking to manage content on the web (blog engines, image database, map interfaces, etc.), and in many cases these or similar technologies may be co-opted for use in systematics. Through mx we seek to adopt the idea of a web-based CMS and make it applicable specifically to systematists.

Existing biodiversity informatics applications tend to focus on discrete data, such as nomenclature (e.g. Page, 2005; Remsen et al., 2006) or specimen management and geo-referencing (e.g. Specify, http://www.specifysoftware.org/Specify). There has been less effort focused at the stages that necessarily occur prior to this discretization - i.e. the formulation of taxonomic descriptions or hypotheses of homology. With its broad flexibility of content classes, the ability to tag (i.e., annotate) data as it is being refined, and use of OTUs at the core of the database mx contains a set of features that will help in this regard.

There are several things that mx is not. At its core mx seeks to be a project-based workbench rather than a central authority. More tightly governed data repositories (e.g., Morphbank, ITIS, uBio, ZooBank (Polaszek, 2005)) should be the final repositories for taxonomic data. A wide range of steps precede the archiving of authoritative data, however, and it is these steps that mx is broadly aimed at managing. While many of mx's content-management functions (e.g., flexible content templates, keyword and image management) are applicable to a wide range of endeavors the supporting functionality is more specifically oriented at providing utilities that practicing

systematists require for day-to-day operations. Mx provides significant mechanisms for recording morphological character data and discussion of results, but it stores neither the results of (e.g. trees) nor the parameters for phylogenetic analyses.

The Rails framework provides various straightforward approaches to allowing for the exchange of data among systems or applications. We illustrate this by providing export formats to Nexus, TNT, and ITIS Identify files and by allowing users to use Morphbank images as if they were mx images. Exporting data is also exemplified by the mapping functionality in mx via the Berkeleymapper interface, which in turn uses Google™ maps.

While development of the type of interconnectivity illustrated by the Berkeleymapper linkages is a priority, it remains secondary to the goal of providing tools that will make process of taxonomy and phylogenetic inference more efficient at their base levels. We believe that many additional steps can be taken in this regard (particularly with respect to work-flow in mx), and that they can develop in parallel with schema-based initiatives (e.g. Taxonomic Concept Schema, http://tdwg.napier.ac.uk/index.php?pagename=TheSchema, see also Kennedy et al., 2006), with the integration of functionality and standard an important long term goal

While many of the ideas that are incorporated into mx have been discussed or developed elsewhere we believe that the combination of features provided by the existing and available application we describe herein represents an relatively infrequently taken approach (thus far) to revisionary systematics. Several thought-experiments can be used in support of this idea (we do not claim that mx provides solutions to these problems, but state them rather as a means of drawing attention to an exciting and rapidly growing field of study):

1) Why are there no accommodations in the ICZN for electronic descriptions of species? If the process of electronic description (i.e. facilitated by a CMS) was widely used this question would be answered.
2) Why does Zootaxa, arguably the most efficient journal for publishing systematics-related (particularly taxonomic) research, require submission by

PDF? If utilities existed to integrate the process of description and phylogenetic inference we would expect to see them seamlessly integrated with publication.

3) Is there any evidence that standardization efforts (e.g. TDWG) have increased the efficiency of taxonomy? Of phylogenetics? Are we seeing more descriptions or hypotheses of phylogeny thanks to standards? While this question is likely unfair given the necessary lag between standards and practices it highlights the need for development of "practical" tools.

With respect to mx, a major project (Table 2.1, Chalcidoid Supermatrix) is underway using mx as the primary means to capture observations and data. Several additional projects presently have existing public front-ends including two taxonomic catalogs and a project which aims to broadly treat wasps of the family Diapriidae. In addition an on-line glossary of terms, based on an ontology constructed in mx is available. The ability to share functionality among different projects in mx has already proved useful. For example Catalogous Evaniidorum (http://evaniid.tamu.edu) and the American Entomological Institute (http://www.amentinst.org) use the same search functionality provided by mx, and three separate projects link back to the glossary of terms. Overall, these projects will continue to provide a critical test of the goals and overall functionality of mx. We anticipate that feedback from these projects and the systematic community in general will help improve subsequent projects and the mx application itself.

Acknowledgements

The mx project is funded by NSF grants DEB-0328922 to Robert Wharton, and DEB-0328920 to Anthony Cognato. Andy Deans and Krishna Dole helped in the preparation of this paper. A number of people have informally discussed issues with Yoder which have directly or indirectly led to mx as it presently exists, these include Bob Wharton, Fred Ronquist, John Oswald, Ed Riley, Zsolte Penzes, Johan Liljeblad,

Norm Johnson, Danny Mandel, Jacques Dubois and members of the BioCorder team. Inspiration has been taken from Rod Page's blog at http://iphylo.blogspot.com/. Bob Wharton provided numerous useful editorial comments. Katja Seltmann and Jiri Hulcr provided significant criticisms and valuable feedback on the mx application.

CHAPTER III

Mannomicrus (HYMENOPTERA: DIAPRIIDAE), A NEW GENUS OF MYRMECOPHILIC DIAPRIID, WITH A DIGITAL VERSION OF MASNER AND GARCÍA'S (2002) KEY TO NEW WORLD DIAPRIINAE AND AN ILLUSTRATION OF DIGITAL DESCRIPTION AND KEY MARKUP USING AN ONTOLOGY[*]

Overview

Mannomicrus Yoder gen. nov. is described with the type species *Hemilexis jessei* Mann, 1914.[1] The genus is only the second myrmecophilic member of the tribe Spilomicrini (Diapriidae: Diapriinae) known from the New World. The new genus is diagnosed versus potentially closely related genera and included in an updated on-line version of a recently published key to New World Diapriinae, which is introduced here. The digital key extends the utility of the original key with additional annotations, navigational functions, and additional images. Both the text of the key and an on-line version of the description can be "marked-up", with words contained there-in checked against an ontology of Hymenoptera morphology terms and linked where matches are found. The usage and means to produce the digital products are briefly reviewed. Both the key and a digital version of the description presented here mark the start of a new web site on diapriid systematics available at http://www.diapriid.org.

---

[*] Reprinted with permission from "*Mannomicrus* (Hymenoptera: Diapriidae), a new genus of myrmecophilic diapriid, with a digital version of Masner and Garcías (2002) key to New World Diapriinae and an illustration of digital description and key markup using an ontology.", by Matthew Yoder, 2007. Zootaxa, 1439, 47-55. Copyright 2007 by Magnolia Press.

[1] None of the taxonomic decisions made herein are to be considered valid under the ICZN, see the originally version in Zootaxa for these purposes.

Introduction

Numerous myrmecophilic diapriids are known for the New World (*e.g.* Huggert and Masner 1983; Loiácono 1981a, 1987, 2000; Loiácono et al. 2000; Loiácono and Margaria 2002; and for a comprehensive list of genera see Masner and García 2002). The vast majority of these genera belong to a single tribe of Diapriinae, the Diapriini, though their association with ants has likely evolved independently in several lineages (Masner, pers. comm., Yoder, unpublished). There are several notable exceptions to the general rule that myrmecophilic species in the New World belong to the Diapriini: 1) some species of *Coecopria*, a genus of uncertain placement (Masner and García 2002), are known to be ant parasites (Loiácono and Margaria 2002); 2) species of *Bruchopria* belong in the Spilomicrini; and 3) the enigmatic *Hemilexis jessei* (Spilomicrini) is reported to be myrmecophilic (Mann 1914). In the Old World *Spilomicrus myrmecophilus* Nixon (Nixon 1947) is the only available record for a myrmecophilic spilomicrine.

Hemilexis jessei Mann, 1914 was last treated in Johnson (1992) where it was transferred, without review, to *Entomacis* based on the synonymy by Muesebeck (1958) of *Hemilexis* Foerster with *Entomacis* Foerster. It is known only from the type series. Yoder (2004), based in part on unpublished information from Lubomir Masner, excluded *H. jessei* from *Entomacis* and left it *incertae-sedis*.

Masner and García (2002) provided a much needed key to the identification of New World Diapriinae, including several newly described genera. As new taxa are discovered, such as the one presented here, it is desirable to extend rather than re-invent Masner and García's (2002) key. Towards this end a digital reproduction of this key is presented here, which adds new functionality, color images, and textual annotations. In addition to the electronic key, this taxon description represents the first published description to be simultaneously made available as an electronic taxon-page using the mx (short for "matrix") content management system (Yoder et al. 2006). The application is available following links at http://www.diapriid.org.

Material and Methods

Descriptions

All known specimens (n=11) of the type series of *Hemilexis jessei* were examined. These specimens are deposited at Harvard University (MCZC: USA, Massachusetts, Cambridge) and the Canadian National Collection of Insects and Arthropods (CNCI: Canada, Ontario, Ottawa). The type series was compared with specimens of all potentially related genera housed at CNCI and the Texas A&M University Insect Collection (TAMU: USA, Texas, College Station). Measurements for the description were taken as in Yoder (2004), and all character states recorded at 60–140x. Terminology follows Yoder (2004) and Masner and García (2002). Descriptive statements are post-fixed with '?' when observations are interpretations based on hidden or very minute characters. These observations need further confirmation via dissection and/or SEM, tasks which were presently impossible given the small type series. All images were taken with a MacroFire camera mounted on a MZ16Apo stereomicroscope and post-processed using AutoMontage™ and Photo-Shop®.

Key, electronic taxon-pages, and ontology

The electronic version of Masner and García's (2002) key and the taxon home page was built using the open-source mx taxonomic content management system described in Yoder et al. (2006). The project's source code and a link to a wiki with further details is available at its Source-Forge® homepage which can be found following hyperlinks at http://www.diapriid.org. Text in the on-line couplets and description are automatically or manually linked to terms stored in the Hymenoptera Glossary (http://hymglossary.tamu.edu, Deans and Yoder 2006). The Hymenoptera Glossary is a simple ontology that provides definitions for and relationships among morphological terms. It its built and managed in an installation of mx. The glossary will ultimately expand into a collaborative effort with members of the Morphbank team, the International Society of Hymenopterists, and the developers of mx.

Taxonomy

Mannomicrus Yoder, *new genus*

Type species : *Hemilexis jessei* Mann, 1914.
Hemilexis jessei Mann, 1914. Original description. Illustrated. Biology.
Hemilexis jessei var. *minor* Mann, 1914.
Entomacis jessei: Johnson, 1992. Cataloged.
Entomacis jessei var. *minor*: Johnson, 1992. Cataloged.
Hemilexis jessei: Yoder, 2004. Considered as Spilomicrini, *incertae sedis*.

Type material: The holotype (deposited at MCZC) of *H. jessei* is in good though somewhat dirty condition, with the left antenna missing segments past the 3rd and the right antenna missing segments past the 11th.

Etymology. A combination of "Mann", in reference to the describer, and "micrus" implying relationship to other spilomicrines. Note that Mann (1914) described *H. jessei* and dedicated it to his "…small collecting companion, Master Jesse Van Law.", it is unclear as to whether Van Law or Mann collected the actual type series.

Classification. *Mannomicrus* is easily recognized as a member of the Spilomicrini by the 13 segmented antennae and characteristic venation (marginal vein relatively long, submarginal clearly separated from anterior margin of forewing). For a further diagnosis of the tribe see Masner and García (2002).

Diagnosis. Most similar to species of the genera *Spilomicrus* and *Bruchopria*, from which it differs characters listed in Table 3.1. *Mannomicrus* can be identified by modifying the key of Masner and 2002) as follows (see also the on-line key and color images available at http://www.diapriid.org):

28 (27). Anterior scutellar pit distinctly bifoveate; basal vein in forewing often present (nebulous); frons unarmed; Nearctic and Neotropical……. ……*Spilomicrus* Westwood [male/female] (part)

- Anterior scutellar pit unifoveate, at most with very slight medial ridge and/or some irregularly spaced longitudinal carinae, or pit absent; basal vein in forewing absent or at most spectral to very slightly sclerotized; frons armed or unarmed ……..……..….. 28a

28a(28). Frons with two sharp points and transverse ledge; body not completely covered with short appressed setae; basal vein in forewing absent; South America [m] (part)…….....…….*Mitropria* Ogloblin

- Frons without two sharp points and transverse ledge (Fig. 3.1, a,b); body completely covered with short appressed setae; basal vein in forewing at most spectral or slightly sclerotized; Mexico [mf]…….. *Mannomicrus* Yoder **New Genus**

TABLE 3.1. Characters diagnosing *Mannomicrus jessei* (Mann) from species of *Spilomicrus* and *Bruchopria*.

Character / Taxon	*Spilomicrus* spp.	*M. jessei* (Mann)	*Bruchopria* spp. *sensu* Masner and García (2002)
anterior scutellar pits	2- prominent clearly separated pits, pits infrequently with scattered carinae	1- carinate throughout, medial most carina sometimes enlarged, but never to the degree found in species of *Spilomicrus*	0- no pit present
posterodorsal pronotum	unmodified, not prominently visible in dorsal view	flat, prominently visible in dorsal view	elevated (dentate), prominently visible in dorsal view
petiole	elongate cylindrical	elongate cylindrical	short, transverse
setae of metasoma/ propodeum	varied, but usually dense and long and never uniform, dense and short	dense and short	dense and short

Description

Female (males unknown), length 2.1–3.0mm, mean=2.6.

Head. Width: 0.46–0.56mm, mean=0.54; height: 0.46–0.61mm, mean=0.66; length: 0.46–0.55mm, mean=0.52 (Fig. 3.1, c); mandible broad, bidentate, with two teeth of subequal length at apex with few scattered erect setae medially and towards base; clypeal ledge reduced, separated from clypeus by narrow groove; clypeus trapezoidal, bearing 4–5 setigerous punctures laterally, setae erect and long; supraclypeal face flat, with all but medial strip densely covered with short, appressed setae, interspersed with few scattered erect longer setae; malar sulcus absent; eye asetose; ommatidia small, surrounded by glabrous patch; dorsal head (frons and post-ocellar vertex), except for small patch surrounding ocellar triangle, densely covered with short appressed setae; posteroventral gena with thick patch of setae, these setae being slightly longer than those appressed on remainder of head; longer erect setae absent on dorsal head and gena; occipital carina complete, short and blade-like; maxillary (3? segmented) and labial (2? segmented) palps very reduced; hypostomal carina short, sharply defined.

Antenna. (Fig. 3.1, f) Scape thickening towards apex, apex with short but well-developed flanges laterally and slight depression ventrally; pedicel reduced, not much larger than A3; A3 slightly longer than or subequal to A4; A4–A7 subequal, cylindrical; A8 broader than A7, following segments subequal; A8–A13 with MGS brush (Yoder 2004), only very slightly flattened ventrally; antennomeres densely covered throughout with short, fine, semi-appressed setae (sensilla); 1–3 erect uniporous? sensilla present along dorsal and ventral surfaces of A8–A13.

FIGURE 3.1. *Mannomicrus jessei* (Mann), female. (a) lateral habitus. (b) dorsal habitus. (c) anterior head.

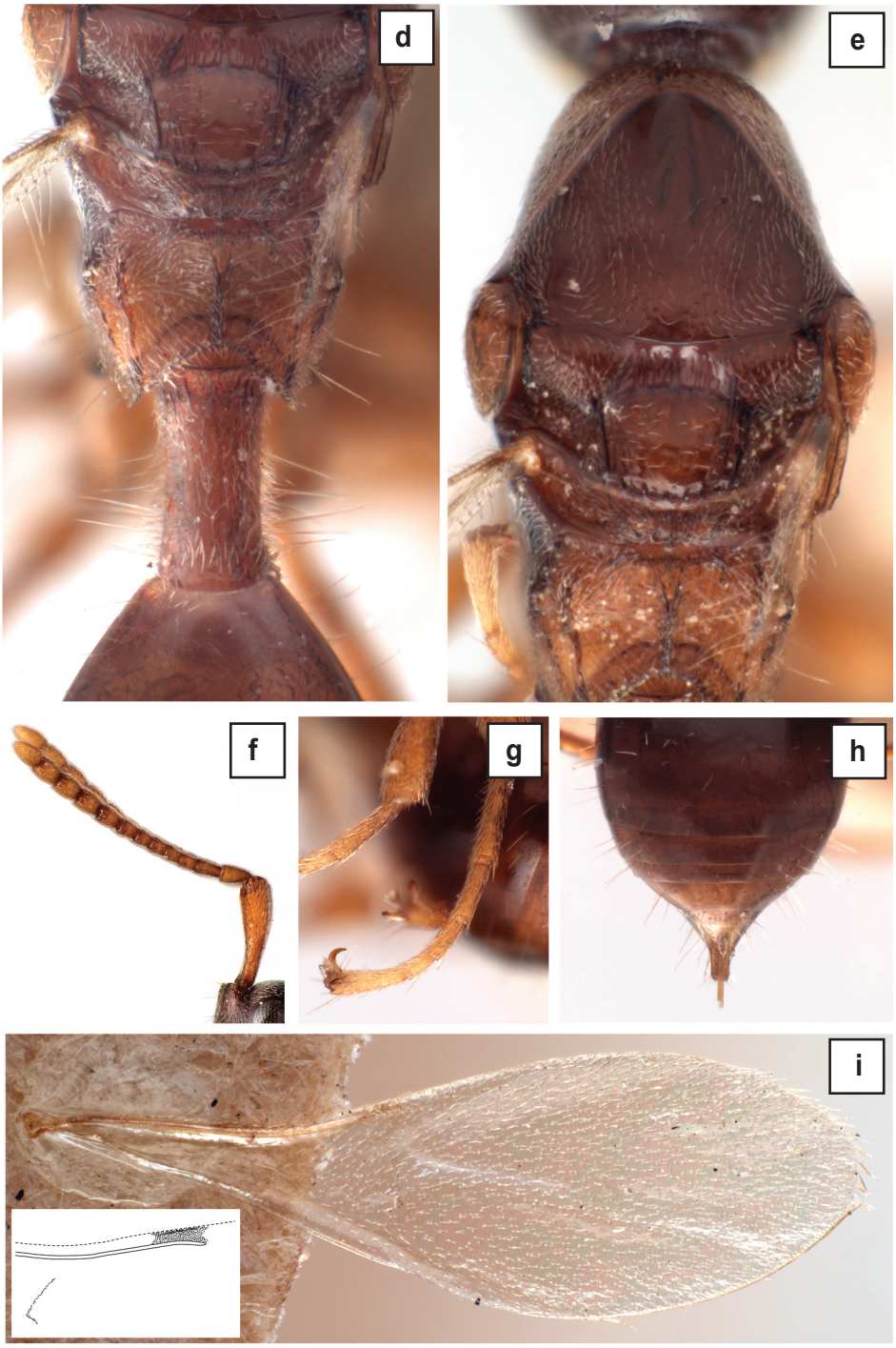

FIGURE 3.1 Continued. (d) scutellum, propodeum, petiole and anterior T2, dorsal view. (e) mesosoma, dorsal view, note that the anterior scutellar pit illustrated here is considered to be unifoveate, see digital key for true bifoveate state. (f) 3.6- antennal, lateral view. (g) 3.7- hind tarsus, note well-developed tarsal claws. (h) posterior metasoma, dorsal view. (i) forewing, note absence of long microtrichia along margin; inset- form of submariginal (=radial) vein junction to anterior margin of forewing.

Mesosoma. Width: 0.50–0.65mm, mean=0.52; height: 0.48–0.63mm, mean=0.57; length: 0.85–1.17mm, mean=1.05 (Figs. 3.1, d,e,g); all of mesosoma, except for small subcircular patch on lateral pronotum and medial strip of dorsal mesoscutum and scutellar disc, densely covered with short, appressed setae, setae of metathorax particularly dense; pro- and mesosoma without any longer semi-erect to erect setae; pronotum in dorsal view (Fig. 3.1, e) broadly visible to near axilla, area anterior to anterior-most mesoscutum unmodified, laterally sub-triangular, only slightly depressed near propleural suture; notauli absent or indicated as fine creases across mesoscutum, mesoscutum where notauli usually located bare (Fig. 3.1, e); anteromedian and parapsidal lines absent or very slightly indicated as shallow depressions; humeral and suprahumeral sulci absent; mesopleuron more or less flat, with few elongate narrow grooves in upper-posterior corner; epicnemial pit very reduced, slit like; median oblique line absent; anterior scutellar pit present, shallow, transversely elongate, with 2–4 longitudinal carinae at its base; posterior junction of axillae to scutellar disc finely carinate; scutellar disc with lateral edges irregularly carinate, lateral-most edge sharply carinate, posterior margin lined with short, irregular scrobiculae; dorsellum with medial keel and lateral keels absent ; axillar depression reduced, not deeply impressed; metapleuron more or less flat, depressed slightly just above hind coxa; outer metanotal process subtriangular, glabrous; propodeum somewhat flattened, carina forming posterior margin very well-developed, sub-horizontal, in dorsal view completely obscuring nuchal area, nucha, and most of petiolar flange; median propodeal keel short (Fig. 3.1, e); all legs relatively uniform in form and as follows- coxa small; trochanter without distinct invagination near apex; femur massive, thickened all but basally, slightly flattened and glabrous ventrally, particularly near apex; tibia with narrow elongate base, widening evenly towards apex; tarsal segment one the longest, two to four subequal, short, five longer (Fig. 3.1, g); pretarsal claws very well-developed, crescent shaped, sharply pointed at apex.

Wings. (Fig. 3.1, i) Forewing submarginal vein separated from anterior margin by clear gap, marginal vein elongate, clearly longer than very short stigmal vein, remaining venation absent to spectral; forewing marginal microtrichia very short (anterior margin) to completely absent (posterior margin); hind wing with only slightly sclerotized

submarginal vein, remaining venation absent, with marginal microtrichia developed similarly to forewing.

Metasoma. (Fig. 3.1, h) Petiole elongate, subcylindrical, with no prominent carinae except near petiolar flange and along ventrolateral most margin, completely densely covered with short appressed to semi-erect setae and with much longer, erect setae on lateral surface; gaster formed by 6 tergites and 5 sternites, with no short appressed setae except for small irregular patch on basoventral S2, with longer, erect setae more or less evenly spaced throughout; ovipositor tip sharp, in some specimens extruded prominently, terminalia otherwise hidden.

Color. Legs, scape, and pedicel light yellow-brown; remaining body brown; anterior and ventral pronotum lighter in some; smaller individuals lighter color overall.

Variation. Mann (1914) was probably led to describe Hemelexis jessei var. minor based on the lighter color andslightly smaller size of the individuals in question. Smaller individuals (including, but not limited to the two specimens labeled Hemelexis jessei var. minor) are generally lighter colored throughout, the pronotum more exposed in dorsal view and the setae appear narrower and slightly lighter. The observed variation is much less than seen in other diapriid species, particularly those that are gregarious parasites, and it is clear that the specimens represent a single species.

Biology. The type series is associated with several ant specimens that Mann (1914) collected. Mann (1914) states these ants to be *Formica subcyanea* Wheeler. This identification is confirmed by a determination attached to the specimen made by T.P. Nuhn 2001, and by a subsequent determination made by myself. Mann (1914) noted that only one or two wasps were present in each colony and that they moved slowly (and freely) among the ants. Masner and García (2002) observed that many specimens of *Bruchopria* have their wings torn or completely removed (likely by ants); this was also the case of most of the specimens of *M. jessei* examined. In some individuals the ovipositor (not sheath) is extruded to a degree not typically seen in other genera of

Diapriinae, and its tip appears quite sharp. The form of the ovipositor would seem to indicate that *M. jessei* is endo- rather than ectoparasitic, though this remains to be proven.

Distribution. All labels bear the same information, "GuerreroMill. | Hidalgo, Mexico | W.M. Mann.", though some are alternately spelled "Guerrer*e*oMill". Mann (1914) notes the locality as "Guerrero Mill,located below Real del Monte, at the Hacienda de Velasco". Guerrero Mill, Mineral del Monte, is at 20.15667N, -98.66W, elevation 2600m, in the state of Hidalgo.

Remarks. While *Mannomicrus jessei* clearly has affinities to species of *Spilomicrus* extending the generic concept of *Spilomicrus*, a very specious genus (> 160 species), to incorporate *M. jessei* is undesirable for several reasons: 1) it would greatly weaken the differential diagnosis of *Spilomicrus* by allowing for an exception to the otherwise uniform characters of form of the anterior scutellar pit and absence of short, appressed pilosity; and 2) it would overlook differences in biology, as species of *Spilomicrus*, with the possible exception of *S. myrmecophilus*, are not known to be myrmecophilic. Based on the description of Nixon (1947) *S. myrmecophilus* shares several similarities with *M. jessei*: 1) the anterior scutellar pit is not paired, being highly reduced; 2) the pronotum is broadly visible in dorsal view; and 3) the ovipositor is long and sharply pointed. However, *S. myrmecophilus* does not exhibit short appressed setae as seen in *M. jessei*. Similar problems with the generic placement of *S. myrmecophilus* exist. Nixon (1947) notes: "I have placed this species in *Spilomicrus* Westwood for reasons of convenience. It cannot be said rightly to belong here. …". As gross morphological convergence is frequently associated with myrmecophily, it will require further study, perhaps molecular, to determine the precise relationships of *M. jessei* and *S. myrmecophilus* to other spilomicrines. The disparate distribution of the two species (Mexico, Mauritius) suggests that similarities in morphology may be due to convergence.

The generic description provided here is relatively specific with respect to setal characteristics. We expect that additional species in this genus, if discovered, may have some variation with respect to the pilosity patterns noted here.

On-line Key to New World Diapriinae

The genus *Mannomicrus* is keyed in an electronic web-accessible version of Masner and García's (2002) key (the English version only at present). The key contains the original illustrations of Masner and García (2002) and the verbatim figure references have been retained. References to new figures are highlighted in the text, as are new text annotations clarifying certain parts. Images should not be assumed to represent all possible morphological forms, as many subtle variations may occur. Determinations should always be confirmed by reference to the more extensive generic descriptions available in Masner and García (2002). The key is available at http://www.diapriid.org.

Discussion

This paper recognizes *Mannomicrus jessei* as only the second truly myrmecophilic member of New World Spilomicrini. The genus *Mannomicrus* is further remarkable for its distribution, as it is currently the only genus of Diapriinae restricted to Mesoamerica (Masner and García 2002). Morphologically *Mannomicrus* is most notable for the short, appressed pilosity and lack of microtrichia along the anterior margin of the forewing, both convergent adaptations found in other myrmecophilic diapriids (Masner and García 2002). These features may be overlooked in the gross examination of bulk samples, so careful examination of additional material will be necessary to identify further specimens should they exist.

The on-line bifurcating key and taxon page were created in an installation of mx. The mx project seeks to provide a wide range of utilities to practicing systematists, one such utility being the creation and management of keys. The interface to mx is completely web based and allows for multiple projects each with multiple users, allowing for collaborations between researchers anywhere in the world. Construction of bifurcating keys in mx is straightforward. At minimum the user fills out either side of a couplet, clicks once to save the data then clicks a button on a given side to add couplets below. The process is continued until the key is complete. New couplets can be inserted or deleted at any point wherein only one side of the present couplet can be presently

followed. Couplets can be figured using images stored in mx (provided by the user) or Morphbank (http://www.morphbank.net). The functionality for using either image type is identical. Each figure can be uniquely annotated. Whole keys (including references to figures) may be duplicated, allowing for sequential updates while retaining original copies.

The electronic key, in addition to being easy to navigate, provides the end-user with several features not available in paper keys: at each couplet the user is shown their decision history (prior couplets), and remaining possible outcomes. Both of these lists contain links, so the user can jump back to a prior couplet or forward to one of the endpoints. Public accessibility of a given key is accomplished by selecting an "is public" option. Because the system is inherently dichotomous, it does not overcome the various problems addressed by interactive keys (see Dallwitz et al. 2005 for an overview), but it is well suited for extension and archiving of historically important dichotomous keys. However, users interested in multi-entry key development should note that mx also provides a multi-entry key engine and matrix manager.

Both the text of the key and the taxonomic description are linkable to a collaboratively built glossary of terms used in Hymenoptera morphology (also built in mx). This linkage greatly extends the usefulness of the text, as definitions can be applied to each matching word. However, some caution is needed when interpreting definitions in this manner. Definitions that are provided in the Hymenoptera glossary may not be those that the author intends. While this is a potential problem, it is also a means to highlight existing problems with, and encourage the adoption of, a shared terminology. The linking functionality is perhaps of most use to those actively writing descriptions and couplets, as their text can be iteratively proofed against a "standard".

It is hoped that new genera for the New World (a minority of which remain to be treated for Diapriinae) will be added to the key as they are published, and that world-wide taxa will ultimately be treated. Because of its digital nature, the on-line key can rapidly be improved with images and textual clarification. To this end, the author welcomes contributions, clarifications, and suggestions for improving the key.

Acknowledgements

Thanks to Bill Wells (TAMU) for specimen imaging, Krishna Dole (TAMU) for help with the development of mx. Robert Wharton, Andy Deans, Matt Buffington, Lubomir Masner and two anonymous reviewers provided editorial comments and insights that greatly improved the paper. Morphbank is hosted at the Florida State University School of Computational Science, Tallahassee, FL 32306-4026 USA. This paper was funded by NSF PEET grant DEB-0328922.

CHAPTER IV

HOSTS OF THE FAMILY DIAPRIIDAE (HYMENOPTERA): A COLLABORATIVE ON-LINE DATABASE OF DIAPRIID BIOLOGICAL RELATIONSHIPS

Overview

Diapriid (Hymenoptera: Diapriidae) biologies are very poorly known, and their supporting literature has never been unified into a single resource. We here introduce an on-line database which seeks to catalog every known diapriid host record, previously published or otherwise. The database currently holds over 450 association records, for over 500 named taxa (parasitoids and hosts), and over 180 references. It is capable of creating complex biological relationships among taxa, which can be linked to various levels of supporting evidence. Each unit of evidence can be arbitrarily ranked as to the confidence level to which it supports the underlying relationship. The database has the capability to track host records attached to individual specimens or lots, and we here record several new host records, previously unreported for the family. Diapriids are found to be primarily endoparasitoids of Diptera emerging from the host pupa. Several records of parasitism of Hymenoptera and Coleoptera are know as well, and those records for Lepidoptera are refuted. Species of several purportedly independent lineages are associated with ants with varying degrees of specificity. Host relationships for identified species are found for only around 150 of the over 2000 described species, or less than 4% of the estimated total diapriid species (Masner, 1993).

Introduction

Diapriids (Hymenoptera: Apocrita: Diapriidae) are a cosmopolitan group of small (< 2mm) parasitic wasps (Masner, 1993). There are just over 2000 available names for the over 4000 species estimated to exist (Masner, 1993). The generic classification for three of the four subfamilies is relatively well resolved, the exception being the subfamily Belytinae. As with other groups of micro-Hymenoptera a high proportion of species have no associated biological information. What is known indicates that diapriids are primarily solitary endo-parasitoids of cyclorrhaphous and nematocerous Diptera, with some species also attacking Hymenoptera (ants, dryinids) and much more rarely Coleoptera (Staphylinidae, Psephenidae).

This paper's focus is on assembling specific records of parasitism (host/parasitoid relationships), and further reference to "host records" herein is specific to this context. Information on other associations (e.g. inquilines, myrmecophily) is also recorded, but this is not the focus of the present work. Much of the information on diapriid biological relationships is secondarily included in works focused on taxonomy or other questions. Identifying the precise mode of parasitism in these cases may not be the primary objective. Caution is thus needed in the interpretation of reported results. Questions such as whether the precise host stage attacked has been identified, whether or not the host was properly isolated and true host remains identified, and from what stage the parasitoid emerged are critical in the circumscription of a parasitoid's biology.

Diapriid host records are typically published singly or in small numbers. A lesser number of publications have aggregated diapriid host records for various purposes (e.g. related to a taxonomic revision, biological control surveys). These works notably include Silvestrii (1914), Thomson (1955), Yasumatsu (1964), Clausen et al. (1965), Teodorescu and Ursa (1979), Muesebeck (1980), and Loiácono and Margaria (2002). There have been no publications however, specifically oriented at unifying the biological literature for species of the family Diapriidae. Diapriid host records have perhaps not been unified because there are apparently relatively few recorded. Several hypotheses as to why diapriid host records are sparse relative to other apocritan groups can be put

forward. Diapriids are not widely used or considered to be useful as biological control agents (though this may not reflect their true potential value as agents). Diapriid immatures have cryptic habits, with hosts typically pupae in areas of rich organic matter, this makes discovery of host bodies extremely difficult. Finally, with the exception of the European fauna, several genera treated for North America, and certain Australian taxa, diapriid systematics essentially operates at the genus group level or higher. The lack of well documented host records is clearly a reflection of the paucity of adequately described species.

This paper introduces a new project, which seeks to unify diapriid host records and describes the initial results of this effort. The results herein are based primarily on literature searches, with the primary goal to identify the majority of host records (i.e. specific records of parasitism, not general biological associations) in the literature. A novel web-based interface has allowed the project to become an international collaboration. We conclude the paper with an outline of the future directions the project will follow.

Materials and Methods

Database

The Diapriid Hosts database is managed by the "associations" component of the 'mx' content-management platform for systematists (see description in Chapter II). It is built using the Ruby on Rails web-application interface on an underlying MySQL database engine. The interface is completely web-based, so data entry is possible anywhere the world-wide web is accessible. A single implementation of 'mx' supports multiple projects each with multiple users. The database's associations (biological) data-model (summarized in Fig. 4.1) allows for taxonomic concepts to be stored as OTUs (operational taxonomic units) independently of governed (e.g. ICZN) names. This allows for verbatim published records to be recorded, and subsequently annotated with the currently accepted name for the taxon in question. Data can be viewed and queried

by a number of different means including taxon name, OTU name, chronologically, by reference, by confidence level, by association type and by host/parasitoid classification.

The database allows for linear relationships of any length (number of taxa) to be recorded. For example "Species A parasitoid of Species B found near Species C". Relationships are defined as a contrasting pair (e.g. parasitoid/host) inserted between the taxa (e.g. Table 4.1). To ease creation of records, once the relationship is defined the taxa may be bound to it in either direction (e.g. A to B or B to A). Any number of associations and association types can be arbitrarily defined.

An important consideration for purported host records is the level of support for that record. After an association is defined it can be tied to supporting evidence, including references, individual specimens, or lots (vouchers), all of which are managed to varying degrees in the mx database. Each record of supporting evidence must be tagged with a level of confidence (e.g. Table 4.2), an arbitrary number of which can be defined and subjectively ordered. Each record of supporting evidence can either positively or negatively support the association (i.e. evidence for the association can be recorded as well as evidence against the relationship).

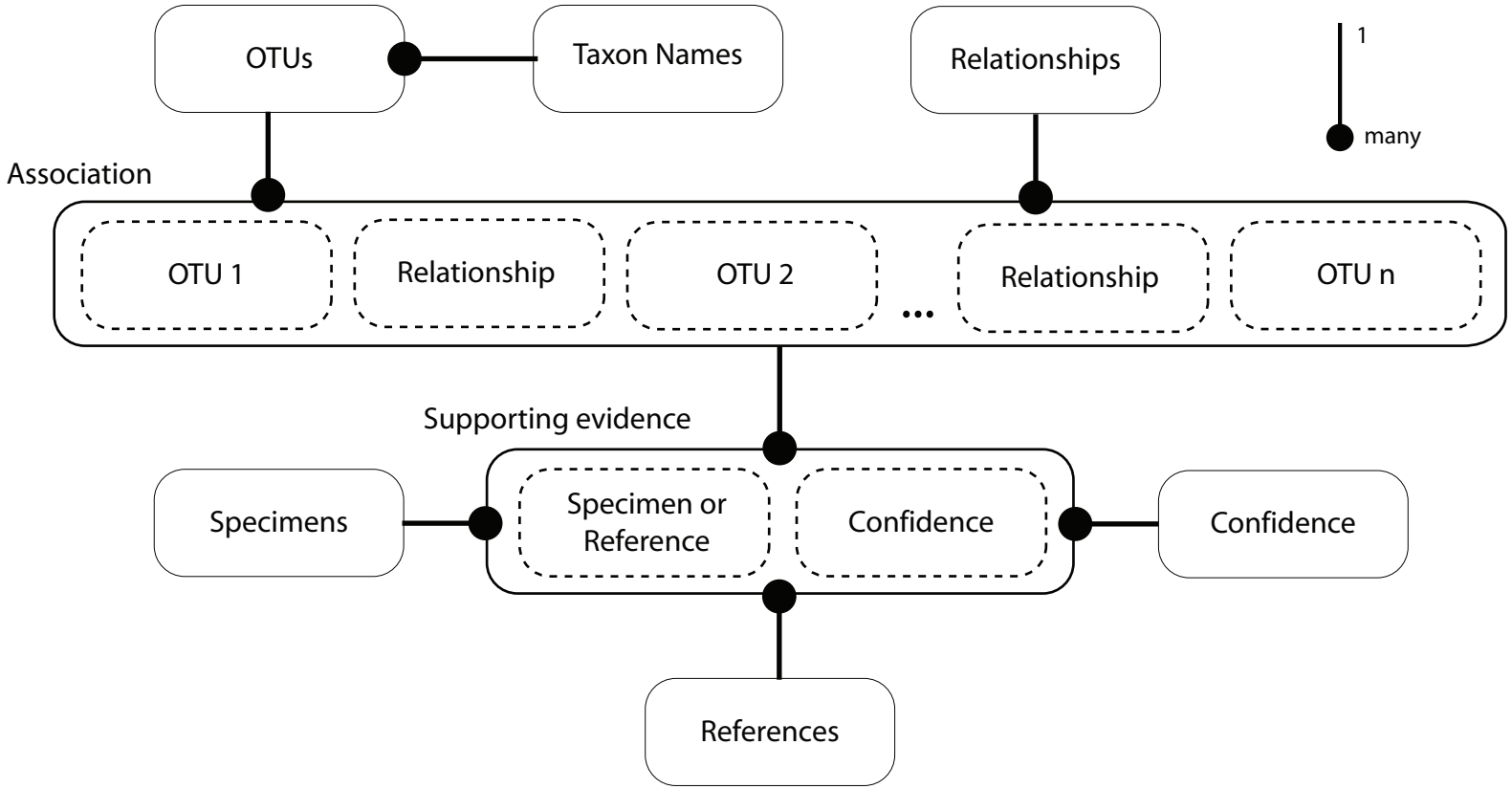

FIGURE 4.1. Simplified relationships implemented in mx used to store host records. Solid shape outlines represent data classes, dashed lines represent class instances. Associations are built by combining, in a linear manner, alternating OTUs and relationships. Each association can be supported (positively or negatively) by many instances of supporting evidence. Supporting evidence is based on a published reference, or a reference to a specimen or group of specimens. OTUs are defined based on the verbatim text as cited in a given reference or specimen determination. They are more precisely identified, when possible, by a taxon name. Qualifiers (notes) can be attached to any data class. Solid circles indicate the 'many' side of a one-to-many relationship, however join tables are not illustrated here.

Data

Literature search. As of submission of this paper records from over 180 references have been added to the database. These were found in searches of larger literature databases (ISI's Web of Science, AGRICOLA, Google Scholar ©), personal literature collections, and by word of mouth. Records listed in Thomson (1955) that made reference to reports in the Review of Applied Entomology (Series A and B, articles ranging from 1916-1936) were further traced to their original citation and confirmed or refuted when possible. A further review of articles within the Review of Applied Entomology under those years was not made. While the database currently contains some references which contain biological but non-host related information, these have not been further parsed in the present version.

Specimen data. While the core data presently available is primarily based on the literature, around 70 associations are based directly on data taken from museum specimens, in particular those from the Canadian National Collection of Insects (CNC).

Relationship meta-data. To reflect the level of certainty and information provided in the reference and specimen data over 20 relationship types (Table 4.1) and 10 confidence levels (Table 4.2) were defined. Several of the relationship types reflect secondary or tertiary relationships of the host to its host plant (e.g. feeding on fruit/host), but a majority reflect the various combinations of endo- and ectoparasitoid, life stage attacked, and solitary or gregarious lifestyles. As idiobiont and koinobiont characters where nearly never included in the literature they were not included in the relationship types, but rather as notes accompanying the association records (presently available on-line).

TABLE 4.1. Association type and use total. A "?" indicates that the start or end of parasitism (i.e. the attacked or emerged-from stage) is unknown. This table is actively updated, and the most recent version can be found at http://www.diapriid.org/public/association/browse_isas.

Interaction	Complement	Total
solitary endoparasitoid	larval host	11
solitary endoparasitoid	pupal host	15
solitary endoparasite	larval-pupal host	3
solitary endoparasitoid	? - pupal	9
solitary endoparasitoid	host	7
solitary ectoparasitoid	host	3
solitary parasitoid	? - pupal	6
gregarious endoparasitoid	larval-pupal host	1
gregarious endoparasitoid	pupal host	7
gregarious endoparasitoid	host	1
gregarious endoparasitoid	larval host	3
super endoparasitoid	pupal host	1
gregarious or super	? - pupal	25
endoparasitoid	larval host	15
endoparasitoid	host	26
endoparasitoid	larval-pupal host	2
endoparasitoid	? - pupal	91
endoparasitoid	pupal host	77
associated with	potential host	46
feeding on tissue	host	2
feeding on fruit	host	27
found in soil near	associated with	1
feeding on leaves	host	5
host	parasitoid	115
predator	prey	1
on	host	10
parasitoid	? - pupal	9
found with	associated with	4

TABLE 4.2. Confidence rankings used in conjunction with supporting evidence, and the number of associations at that rank. Confidence level decreases towards bottom of the table. This table is actively updated, and the most recent version can be found at http://www.diapriid.org/public/association/browse_confidences.

Confidence	Total
Multiple rearings from known lab monoculture.	42
Multiple isolated single hosts from wild, with host remains isolated following emergence/dissection.	172
Isolated single host, with host remains isolated following emergence/dissection.	23
Single parasite/parasitoid from culture.	1
Single specimen mounted with purported host remains attached, no further information.	2
Cited with no further explanation (e.g. in table/list).	171
Note attached to pinned specimen/lot, no host remains.	2
Museum specimen(s) and labels, with no further data.	78
Multiple specimens in collection with single unconfirmed host/host remains included, no further data.	1
Word of mouth, no reference to specimens.	1
Unchecked/unclassified citation.	13

Taxonomic names. All names were recorded as presented in the literature or on specimen labels. This was done by creating an Operational Taxonomic Unit (OTU) in mx with the provided name. OTUs where then tagged with the currently valid taxonomic name when discernable, when not a note attached to the OTU record. The majority of accepted nomenclature follows two sources, Johnson (1992) (also available on-line via the Hymenoptera On-Line Database at http://iris.biosci.ohio-state.edu/hymenoptera/) and the BioSystematic Database of World Diptera (Thompson, 2005).

Results and Discussion

Since they are extensively linked and various customizable reports are available results are best viewed on-line (http://www.diapriid.org/ following the 'associations' link). Over 450 associations were recorded, but of these only around 150 include a reference to a parasitoid species name. This suggests that host records exist for less than 4% of the estimated total number of diapriid species. Of the total associations the large majority of those are for the subfamily Diapriinae, and within this subfamily most belong to species of the genera *Trichopria* (Diapriini) and *Coptera* (Psilini). On a per taxon level most host records are attributed to unidentified or undescribed species of *Trichopria* (over 50 associations). This is not unexpected as of the diapriids *Trichopria* is likely the most speciose genus of diapriids, and also one of the most morphologically uniform and thus taxonomically problematic. Fewer than 15 unique records are recorded for the Belytinae, all from nematocerous Diptera with the exception of single records for species of *Synacra* on *Musca domestica* (Float et al., 1999, 2002) and a ponerine ant (Nixon, 1957). The Ismarinae contain a single genus that is both morphologically and biologically isolated from the remaining Diapriidae. They are exclusively hyperparasitoids of Auchenorrhyncha (Hemiptera) through dryinids (Hymenoptera: Dyrinidae) (Chambers, 1955, 1981; Jervis 1979).

Perhaps most remarkable, as noted elsewhere (Masner, 1993), only a single host record exists for the Ambositrinae, and this was recorded over 100 years ago (Hudson, 1892, Marshal 1892). This lack of evidence comes in spite of modern generic revisions of the Australasian fauna by Naumann (1982, 1987, 1988). Since ambositrines are extremely frequently encountered in Valdivian South America, New Zealand and Australia, concentrated efforts at discovering their hosts should be initiated there. Based on their highly varied morphology (e.g. some Australian species *Acanthobetyla* are remarkably ant-like in form while many Chilean species exhibit a habitus more typically seen in purportedly plesiomorphic diapriids) the range of hosts attacked by members of this subfamily may be diverse.

TABLE 4.3. Current (March, 2007) relationship data in the diapriid host-parasite database. The data can be sorted by various criteria, and in many cases notes or additional clarifications are available on-line at http://www.diapriid.org/, following the link to "associations". See on-line database for further information on "Specimen" in the "Citation" column, this refers to data captured from specimens which may or may not have been previously published. Relationships without citations are primarily specimen-based, where specimens records have not been formally captured to the database, see on-line version for notes in this case. Data are sorted by parasitoid, then host-family, then relationship type. See Table 4.1 for further information on relationship types.

Parasitoid	[Relationship] / Host(s)	Citation
Diapriidae	[endoparasitoid / ? - pupal] *Drosophila* (Drosophilidae)	Clausen et al. (1965)
	[endoparasitoid / ? - pupal] *Fannia femoralis* (Fanniidae)	Legner and Olton (1971)
	[endoparasitoid / pupal host] *Fannia canicularis* (Fanniidae)	Legner and Olton (1971)
	[endoparasitoid / pupal host] *Musca domestica* (Muscidae)	Legner and Olton (1971)
	[solitary endoparasitoid / host] *Musca domestica* (Muscidae)	Hogsette et al. (2001)
	[solitary endoparasitoid / host] *Musca domestica* (Muscidae)	Hogsette et al. (2001)
	[gregarious endoparasitoid / pupal host] *Inopus rubriceps* (Stratiomyidae)	Robertson (1987)
	[parasitoid / host] *Carcelia evolans* (Tachinidae) [on / host] *Imbrasia cytherea*	van den Berg (1974)
	[endoparasitoid / ? - pupal] Tephritidae	Clausen et al. (1965)
Ambositrinae		
Betyla fulva	[solitary endoparasitoid / ? - pupal] *Arachnocampa luminosa* (Keroplatidae)	Marshall (1892); Hudson (1892)
Belytinae		
Belytinae	[endoparasitoid / larval host] Sciaridae	Nixon (1957)
Aclista	[host / parasitoid] Mycetophilidae	
Acropiesta flaviventris	[endoparasitoid / larval host] *Trichosia* (Sciaridae)	Huggert, L. (1979)
Cinetus lanceolatus	[endoparasitoid / larval host] Mycetophilidae [feeding on tissue / host] *Boletus*	Nixon (1957)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Eumiota longiventris	[endoparasitoid / larval host] Mycetophilidae [feeding on tissue / host] *Suillus variegatus*	Huggert, L. (1979)
Stylaclista quasimodo	[solitary endoparasitoid / ? - pupal] Cecidomyiidae	Early (1980)
Synacra	[endoparasitoid / pupal host] *Musca domestica* (Muscidae)	Floate et al. (1999); Floate et al. (2002)
	[solitary endoparasitoid / larval host] *Bradysia difformis* (Sciaridae)	Hellqvist (1994)
Synacra brachialis	[endoparasitoid / host] *Ponera coarctata* (Formicidae)	Nixon (1957)
Synacra paupera	[endoparasitoid / host] *Bradysia difformis* (Sciaridae)	Notton (1997)
Synacra sociabilis	[associated with / found with] *Formica* (Formicidae)	
Diapriinae		
Diapriinae	[solitary parasitoid / ? - pupal] *Helosciomyza subalpina* (Helosciomyzidae)	Early and Horning (1978)
	[endoparasitoid / host] *Haematobia irritans* (Muscidae)	Thomas and Morgan (1972)
	[endoparasitoid / ? - pupal] *Inopus rubriceps* (Stratiomyidae)	Osborn (1974)
	[gregarious endoparasitoid / larval-pupal host] *Tabanus nigrovittatus* (Tabanidae)	Magnarelli and Anderson (1980)
Spilomicrini	[parasitoid / host] *Austrothaumalea denticulata* (Thaumaleidae)	Sinclair (2000)
Abothropria lloydi	[parasitoid / host] *Glossina palpalis* (Glossinidae)	Thompson (1955)
Acanthopria	[associated with / potential host] *Cyphomyrmex* (Formicidae)	Specimen
	[associated with / potential host] *Neivamyrmex gibbatus* (Formicidae)	Specimen
	[solitary endoparasitoid / larval host] *Cyphomyrmex minutus* (Formicidae)	Fernández-Marín et al. (2006)
	[solitary endoparasitoid / larval host] *Cyphomyrmex rimosus* (Formicidae)	Fernández-Marín et al. (2006)
Acanthopria gracilicornis	[parasitoid / host] *Eciton burchelli* (Formicidae)	Loiácono and Margaria (2002)
	[parasitoid / host] *Eciton quadriglume* (Formicidae)	Loiácono and Margaria (2002)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Acanthopria myrmecophila	[parasitoid / host] *Eciton quadriglume* (Formicidae)	Loiácono and Margaria (2002)
Acanthopria rudebecki	[parasitoid / host] *Lepisiota* (Formicidae)	
Aneurhynchus	[found with / associated with] mushrooms	Specimen
Aneurhynchus fannivorus	[endoparasitoid / ? - pupal] *Fannia* (Fanniidae)	Honda (1968)
Aneuropria foersteri	[solitary parasitoid / ? - pupal] *Piophila casei* (Piophilidae)	Teodorescu and Ursu (1979)
	[parasitoid / host] *Rhagoletis cerasi* (Tephritidae)	Thompson (1955)
Antarctopria coelopae	[gregarious or super / ? - pupal] *Calliphora* (Calliphoridae)	Early (1978)
	[gregarious or super / ? - pupal] *Calliphora erythrocephala* (Calliphoridae)	Early (1978)
	[gregarious or super / ? - pupal] *Baeopterus philpotti* (Coelopidae)	Early (1978)
	[gregarious or super / ? - pupal] *Baeopterus robustus* (Coelopidae)	Early (1978)
	[gregarious or super / ? - pupal] *Chaetocoelopa littoralis* (Coelopidae)	Early (1978)
	[gregarious or super / ? - pupal] *Icaridion debile* (Coelopidae)	Early (1978)
Antarctopria diomedeae	[solitary endoparasitoid / pupal host] *Leptocera* (Sphaeroceridae)	Early (1980)
	[solitary endoparasitoid / pupal host] *Apetaenus littorea* (Tethinidae)	Early (1978)
Antarctopria latigaster	[endoparasitoid / larval-pupal host] Coelopidae	Gressitt et al. (1964)
	[solitary endoparasitoid / ? - pupal] *Icaridion debile* (Coelopidae)	Early (1978)
	[solitary endoparasitoid / pupal host] Syrphidae	
	[solitary endoparasitoid / ? - pupal] *Apetaenus australis* (Tethinidae)	Early (1980)
Apopria	[associated with / potential host] *Neivamyrmex carolinensus* (Formicidae)	Specimen
	[associated with / potential host] *Neivamyrmex opacithorax* (Formicidae)	
Apopria coveri	[associated with / potential host] *Neivamyrmex texanus* (Formicidae)	Specimen
	[potential host / associated with] *Neivamyrmex opacithorax* (Formicidae)	Masner and Garcia (2002)
Asolenopsia	[associated with / potential host] Formicidae	Specimen
Asolenopsia mutilata	[associated with / potential host] *Neivamyrmex legionis* (Formicidae)	Specimen

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[parasitoid / host] *Neivamyrmex legionis* (Formicidae)	Loiácono and Margaria (2002)
	[parasitoid / host] *Neivamyrmex pseudops* (Formicidae)	Loiácono and Margaria (2002)
Asolenopsia schwarzmaieri	[associated with / potential host] Formicidae	Specimen
	[parasitoid / host] *Neivamyrmex pseudops* (Formicidae)	Loiácono and Margaria (2002)
Auxopaedeutes	[associated with / potential host] Formicidae	Specimen
	[associated with / potential host] *Solenopsis molesta* (Formicidae)	Specimen
	[associated with / potential host] *Solenopsis truncorum* (Formicidae)	Specimen
	[solitary endoparasitoid / larval host] Formicidae	Specimen
Basalys	[endoparasitoid / ? - pupal] Diptera	Specimen
	[parasitoid / host] Diptera	Masner and Garcia (2002); Simmonds (1952)
	[parasitoid / ? - pupal] Chloropidae	
	[associated with / potential host] *Cryptorhynchus lapathi* (Curculionidae)	Specimen
	[associated with / potential host] Scolytinae (Curculionidae)	Specimen
	[endoparasitoid / host] *Ceutorhynchus assimilis* (Curculionidae)	Specimen
	[parasitoid / host] *Spiniphora dorsalis* (Phoridae)	Specimen
	[parasitoid / ? - pupal] Phoridae	
	[solitary endoparasitoid / pupal host] *Megaselia agarici* (Phoridae) [feeding on fruit / host] *Agaricus augustus*	Notton (1991)
	[solitary endoparasitoid / larval host] *Psila rosae* (Psilidae)	Specimen
Basalys parva	[solitary endoparasitoid / pupal host] *Megaselia agarici* (Phoridae) [feeding on fruit / host] *Agaricus augustus*	Notton (1991)
Basalys pegomyiae	[parasitoid / host] *Pegomya* (Anthomyiidae)	Brues (1908)
Basalys tritoma	[solitary endoparasitoid / ? - pupal] *Dacnusa* (Braconidae) [solitary endoparasitoid / ? - pupal] *Psila rosae* (Psilidae)	Wright et al. (1947)
	[endoparasitoid / ? - pupal] *Oscinella frit* (Chloropidae)	Imms (1930)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[parasitoid / host] *Oscinella frit* (Chloropidae)	Thompson (1955)
	[parasitoid / host] *Thaumatomyia notata* (Chloropidae)	Thompson (1955)
	[solitary endoparasitoid / pupal host] *Psila rosae* (Psilidae)	Wright et al. (1947)
Bruchopria	[associated with / potential host] *Solenopsis richteri* (Formicidae)	Masner and Garcia (2002); Specimen
Bruchopria	[endoparasitoid / ? - pupal] *Paratheresia claripalpis* (Tachinidae)	Masner and Garcia (2002)
Bruchopria hexatoma	[parasitoid / host] *Solenopsis saevissima* (Formicidae)	Loiácono and Margaria (2002)
Bruchopria tucumana	[parasitoid / host] *Sarcophaga* (Sarcophagidae)	Thompson (1955)
	[parasitoid / host] *Lydella* (Tachinidae)	Loiácono and Margaria (2002)
	[parasitoid / host] *Paratheresia claripalpis* (Tachinidae)	Loiácono and Margaria (2002)
	[parasitoid / host] *Paratheresia claripalpis* (Tachinidae)	Thompson (1955)
Bruesopria	[associated with / potential host] Formicidae	Specimen
Coecopria plaumanni	[parasitoid / host] *Camponotus rufipes* (Formicidae)	Loiácono and Margaria (2002)
Coecopria pygmea	[parasitoid / host] *Camponotus rufipes* (Formicidae)	Loiácono and Margaria (2002)
Coptera	[parasitoid / host] Milichiidae	Muesebeck (1980)
	[endoparasitoid / ? - pupal] *Stomoxys calcitrans* (Muscidae)	Hoggsette et al. (1994)
	[endoparasitoid / host] *Musca domestica* (Muscidae)	Specimen
	[parasitoid / host] Muscidae	Muesebeck (1980)
	[parasitoid / host] Otitidae	Muesebeck (1980)
	[parasitoid / host] Psilidae	Muesebeck (1980)
	[endoparasitoid / pupal host] *Anastrepha* (Tephritidae)	Ovruski et al. (2000)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[endoparasitoid / pupal host] *Anastrepha ludens* (Tephritidae)	Baker et al. (1944); McPhail and Bliss (1933); Stibick (2004); Thompson (1955)
	[endoparasitoid / pupal host] *Bactrocera dorsalis* (Tephritidae)	Puttarudriah and Usman (1954)
	[parasitoid / host] *Rhagoletis suavis completa* (Tephritidae)	Thompson (1955)
	[parasitoid / host] Tephritidae	Thompson (1955)
	[solitary endoparasitoid / ? - pupal] Tephritidae [feeding on fruit / host] *Coffea*	Wharton et al. (2000)
Coptera atricornis	[parasitoid / host] *Lonchaea corticis* (Lonchaeidae)	Muesebeck (1980)
	[parasitoid / host] *Pseudotephritis corticalis* (Ulidiidae)	Muesebeck (1980)
Coptera chylizae	[endoparasitoid / ? - pupal] *Chyliza notata* (Psilidae)	Muesebeck (1980)
Coptera cingulatae	[endoparasitoid / ? - pupal] *Rhagoletis cingulata* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / ? - pupal] *Rhagoletis fausta* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / ? - pupal] *Rhagoletis pomonella* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / ? - pupal] *Rhagoletis suavis* (Tephritidae)	Muesebeck (1980)
Coptera evansi	[endoparasitoid / pupal host] *Euphranta* (Rhacochlaena) *canadensis* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / pupal host] *Rhagoletis* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / pupal host] *Rhagoletis completa* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / pupal host] *Rhagoletis fausta* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / pupal host] *Rhagoletis juglandis* (Tephritidae)	Muesebeck (1980)
Coptera gestroi	[solitary parasitoid / ? - pupal] *Lucilia* (Calliphoridae)	Teodorescu and Ursu (1979)
Coptera haywardi	[endoparasitoid / larval host] Braconidae [endoparasitoid / larval-pupal host] *Anastrepha suspensa* (Tephritidae)	Sivinski et al. (1998)
	[? - pupal / endoparasitoid] *Coptera haywardi* (Diapriidae)	Loiacono (1981b)
	[endoparasitoid / ? - pupal] *Anastrepha* (Tephritidae)	Ovruski et al. (2000)
	[endoparasitoid / ? - pupal] *Anastrepha* (Tephritidae) [on / host] *Eugenia uniflora*	Aguiar-Menezes et al. (2003)
	[endoparasitoid / ? - pupal] *Anastrepha fraterculus* Wiedemann (Tephritidae)	Loiacono (1981b)
	[endoparasitoid / ? - pupal] *Anastrepha fraterculus* Wiedemann (Tephritidae) [feeding on fruit / host] *P. guajava*	Sivinski et al. (2000)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[endoparasitoid / ? - pupal] *Anastrepha obliqua* (Tephritidae) [feeding on fruit / host] *Spondias purpurea*	Sivinski et al. (2000)
	[endoparasitoid / ? - pupal] *Anastrepha serpentina* (Tephritidae) [feeding on fruit / host] *Chrysophyllum cainito*	García and Montilla (2001)
	[endoparasitoid / ? - pupal] *Anastrepha striata* (Tephritidae) [feeding on fruit / host] *Spondias mombin*	García and Montilla (2001)
	[endoparasitoid / pupal host] *Anastrepha fraterculus* Wiedemann (Tephritidae) [feeding on fruit / host] *P. guajava*	López et al. (1999)
	[endoparasitoid / pupal host] *Anastrepha ludens* (Tephritidae)	Lopez (1996); Sivinski et al. (1998); Larissa et al. (2002)
	[endoparasitoid / pupal host] *Anastrepha ludens* (Tephritidae) [feeding on fruit / host] *Citrus sinensis*	López et al. (1999); López et al. (1999)
	[endoparasitoid / pupal host] *Anastrepha obliqua* (Tephritidae)	García and Montilla (2001)
	[endoparasitoid / pupal host] *Anastrepha striata* (Tephritidae) [feeding on fruit / host] *P. guajava*	López et al. (1999)
	[endoparasitoid / pupal host] *Anastrepha suspensa* (Tephritidae)	Sivinski et al. (1998)
	[endoparasitoid / pupal host] *Ceratitis capitata* (Tephritidae)	Sivinski et al. (1998); Baeza-Larios et al. (2002)
	[endoparasitoid / pupal host] *Toxotrypana curvicauda* (Tephritidae)	Sivinski et al. (1998)
Coptera muscidorum	[endoparasitoid / ? - pupal] *Glossina palpalis* (Glossinidae)	Nixon (1930)
Coptera occidentalis	[endoparasitoid / ? - pupal] *Rhagoletis cingulata* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / ? - pupal] *Rhagoletis completa* (Tephritidae)	Muesebeck (1980)
	[endoparasitoid / host] *Rhagoletis completa* (Tephritidae)	Specimen
	[endoparasitoid / pupal host] *Ceratitis capitata* (Tephritidae)	Kazimirova et al. (1997); Kazimirova and Ortel (2000)
	[solitary endoparasitoid / pupal host] *Ceratitis capitata* (Tephritidae)	Kazimirova and Vallo (1992)
Coptera pholeomyiae	[associated with / potential host] *Atta texana* (Formicidae)	Specimen
	[endoparasitoid / ? - pupal] *Pholeomyia comans* (Milichiidae) [associated with / potential host] *Atta texana* (Formicidae)	Muesebeck (1980)
Coptera pomonellae	[endoparasitoid / ? - pupal] *Rhagoletis pomonella* (Tephritidae)	Muesebeck (1980)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[solitary endoparasitoid / ? - pupal] *Rhagoletis suavis* (Tephritidae)	Muesebeck (1980)
Coptera punctiger	[parasitoid / host] *Drosophila* (Drosophilidae)	Muesebeck (1980)
Coptera robustior	[endoparasitoid / ? - pupal] *Ceratitis contramedia* (Tephritidae) [feeding on fruit / host] *Waburgia*	Clausen et al. (1965)
	[endoparasitoid / pupal host] *Ceratitis punctata* (Tephritidae) [feeding on leaves / host] *Tabernaemontana longiflora*	Silvestri (1914)
Coptera silvestrii	[parasitoid / host] Eulophidae (Eulophidae) [endoparasitoid / pupal host] *Ceratitis capitata* (Tephritidae)	Pemberton and Willard (1918); Thompson (1955)
	[endoparasitoid / ? - pupal] Tephritidae	Clausen et al. (1965)
	[endoparasitoid / pupal host] *Bactrocera oleae* (Tephritidae)	Silvestri (1914)
	[endoparasitoid / pupal host] *Ceratitis anonae* (Tephritidae) [feeding on fruit / host] *Dovysalis*	Silvestri (1914)
	[endoparasitoid / pupal host] *Ceratitis capitata* (Tephritidae)	Silvestri (1914)
	[endoparasitoid / pupal host] *Ceratitis colae* (Tephritidae)	Silvestri (1914)
	[endoparasitoid / pupal host] *Ceratitis contramedia* (Tephritidae) [feeding on fruit / host] *Waburgia*	Clausen et al. (1965)
	[endoparasitoid / pupal host] *Ceratitis cosyra* (Tephritidae) [feeding on fruit / host] *Chrysobalanus ellipticus*	Silvestri (1914)
	[endoparasitoid / pupal host] *Ceratitis giffardi* (Tephritidae)	Silvestri (1914)
	[endoparasitoid / pupal host] *Ceratitis rosa* (Tephritidae) [feeding on fruit / host] *P. guajava*	Clausen et al. (1965)
	[endoparasitoid / pupal host] *Ceratitis simi* (Tephritidae) [feeding on fruit / host] *Acokanthera schimperi*	Clausen et al. (1965)
	[endoparasitoid / pupal host] Tephritidae	Silvestri (1914)
	[endoparasitoid / pupal host] *Trirhithrum nigerrimum* (Tephritidae) [feeding on fruit / host] *Coffea arabica*	Silvestri (1914)
	[parasitoid / host] *Bactrocera cucurbitae* (Tephritidae)	Fullaway (1918); Thompson (1955)
	[parasitoid / host] *Bactrocera latifrons* (Tephritidae)	Narayanan and Chawla (1962)
	[parasitoid / host] *Bactrocera oleae* (Tephritidae)	Thompson (1955)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[parasitoid / host] *Ceratitis capitata* (Tephritidae)	Thompson (1955)
Coptera strauziae	[endoparasitoid / ? - pupal] *Strauzia longipennis* (Tephritidae) [on / host] *Helianthus tuberosus*	Muesebeck (1980); Specimen
Diapria	[endoparasitoid / ? - pupal] *Eristalis* (Syrphidae)	Specimen
	[endoparasitoid / ? - pupal] Syrphidae	Specimen
	[solitary endoparasitoid / pupal host] *Eristalis* (Syrphidae)	Specimen
Diapria coccophaga	[endoparasitoid / host] *Ctenochiton perforatus*	Gourlay (1930); Thompson (1955)
Diapria conica	[gregarious or super / ? - pupal] *Eristalis tenax* (Syrphidae)	Sanders (1911)
	[parasitoid / host] *Eristalis tenax* (Syrphidae)	Fahringer (1922); Vukasovic (1926); Vukasovic (1928); Thompson (1955);
Diapria solitaria	[parasitoid / host] *Dendrolimus pini*	Thompson (1955)
Doliopria collegii	[parasitoid / host] *Eciton burchelli* (Formicidae)	Loiácono and Margaria (2002)
Doliopria collegii	[parasitoid / host] *Eciton quadriglume* (Formicidae)	Loiácono and Margaria (2002)
Doliopria flavipes	[parasitoid / host] *Eciton burchelli* (Formicidae)	Loiácono and Margaria (2002)
Ecitovagus gibbus	[associated with / potential host] Formicidae	Specimen
	[associated with / potential host] *Neivamyrmex nigrescens* (Formicidae)	Specimen
Entomacis californica	[parasitoid / host] *Forcipomyia texana* (Ceratopogonidae)	Yoder (2004)
Entomacis longii	[endoparasitoid / ? - pupal] *Forcipomyia wheeleri* (Ceratopogonidae)	Ashmead (1902); Muesebeck and Walkley (1951)
Ferrieropria	[on / host] *Lepisiota* (Formicidae)	

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Hemilexomyia abrupta	[endoparasitoid / ? - pupal] *Calliphora stygia* (Calliphoridae)	Dodd (1920)
	[parasitoid / host] *Calliphora stygia* (Calliphoridae)	Thompson (1955)
	[parasitoid / host] *Calliphoridae*	Dodd (1930); Thompson (1955)
	[parasitoid / host] *Pollenia* (Calliphoridae)	Thompson (1955)
	[endoparasitoid / ? - pupal] *Musca domestica* (Muscidae)	Froggatt (1918); Johnston and Tiegs (1922)
	[endoparasitoid / ? - pupal] *Ophyra* (Muscidae)	Froggatt (1917); Dodd (1920); Dodd (1930); Johnston and Tiegs (1922)
	[parasitoid / host] *Musca domestica* (Muscidae)	Thompson (1955)
	[parasitoid / host] *Ophyra* (Muscidae)	Thompson (1955)
Hemilexomyia spinosa	[solitary endoparasite / larval-pupal host] *Limnohelina* (Muscidae)	Early (1980)
Idiotypa	[endoparasitoid / host] Diptera	Specimen
Idiotypa nigriceps	[endoparasitoid / ? - pupal] Phoridae	Masner and Garcia (2002)
Labidopria	[associated with / potential host] *Eciton* (Formicidae)	Specimen
	[associated with / potential host] Formicidae	Specimen
	[associated with / potential host] *Labidus* (Formicidae)	Specimen
	[associated with / potential host] *Labidus praedator* (Formicidae)	Specimen
Labidopria longicornis	[parasitoid / host] *Labidus praedator* (Formicidae)	Loiácono and Margaria (2002)
Leaiopria termitarii	[associated with / potential host] *Nasutitermes fumigatus* (Termitidae)	Specimen
Lepidopria aberrans	[endoparasitoid / ? - pupal] *Cryptomeigenia theutis* (Tachinidae) [on / host] *Phyllophaga inversa*	Brues (1916); Thompson (1955)
Leucopria	[associated with / potential host] *Apterostigma auriculatum* (Formicidae)	Specimen

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Mimopria	[associated with / potential host] *Eciton hamatum* (Formicidae)	Specimen
	[associated with / potential host] *Eciton hamatum* (Formicidae)	Specimen
	[parasitoid / host] *Neivamyrmex goledii* (Formicidae)	Loiácono and Margaria (2002)
Mimopria barbata	[parasitoid / host] *Nomamyrmex esenbeckii* (Formicidae)	Loiácono and Margaria (2002)
Mimopria campbellorum	[associated with / potential host] *Eciton hamatum* (Formicidae)	Specimen
	[parasitoid / host] *Eciton hamatum* (Formicidae)	Loiácono and Margaria (2002)
Mimopria comes	[parasitoid / host] *Labidus coecus* (Formicidae)	Loiácono and Margaria (2002)
	[parasitoid / host] *Nomamyrmex esenbeckii* (Formicidae)	Loiácono and Margaria (2002)
	[parasitoid / host] *Nomamyrmex hartigii* (Formicidae)	Loiácono and Margaria (2002)
Mimopria ecitophila	[associated with / potential host] *Eciton hamatum* (Formicidae)	Specimen
	[parasitoid / host] *Eciton hamatum* (Formicidae)	Loiácono and Margaria (2002)
Mimopriella	[associated with / potential host] *Labidus coecus* (Formicidae)	Specimen
	[associated with / potential host] *Trachymyrmex* (Formicidae)	Specimen
	[solitary endoparasitoid / larval host] *Cyphomyrmex rimosus* (Formicidae)	Fernández-Marín et al. (2006)
Mimopriella pentatoma	[parasitoid / host] *Neivamyrmex goledii* (Formicidae)	Loiácono and Margaria (2002)
Monelata	[solitary endoparasitoid / ? - pupal] *Atherigona soccata* (Muscidae)	Taley and Thakare (1979)
Monelata parvula	[gregarious or super / ? - pupal] *Lucilia* (Calliphoridae)	Teodorescu and Ursu (1979)
Myrmecopria mellea	[solitary endoparasitoid / larval host] *Neivamyrmex carolinensus* (Formicidae)	Specimen
	[solitary endoparasitoid / larval host] *Neivamyrmex opacithorax* (Formicidae)	Specimen

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Neivapria penicillata	[parasitoid / host] *Neivamyrmex minenses* (Formicidae)	Loiácono and Margaria (2002)
Neurogalesus	[solitary endoparasitoid / pupal host] *Inopus rubriceps* (Stratiomyidae)	Robertson and Zalucki (1985); Robertson (1987)
Neurogalesus militis	[endoparasitoid / ? - pupal] *Inopus rubriceps* (Stratiomyidae)	Osborn (1974); Robertson (1987)
	[endoparasitoid / ? - pupal] *Inopus rubriceps* (Stratiomyidae)	Osborn et al. (1973)
	[solitary endoparasitoid / pupal host] *Inopus rubriceps* (Stratiomyidae)	Robertson and Zalucki (1985)
Notoxoides	[associated with / potential host] Formicidae	Specimen
	[associated with / potential host] Formicidae	Specimen
	[associated with / potential host] Formicidae	Specimen
	[associated with / potential host] *Neivamyrmex cristatus* (Formicidae)	Specimen
Notoxoides cornutus	[associated with / potential host] *Neivamyrmex cristatus* (Formicidae)	Specimen
Notoxoides pedissequus	[parasitoid / host] *Neivamyrmex pseudops* (Formicidae)	Loiácono and Margaria (2002)
Notoxoides pronotalis	[parasitoid / host] *Eciton dulcium* (Formicidae)	Loiácono and Margaria (2002)
Oxypria collegiales	[parasitoid / host] *Eciton quadriglume* (Formicidae)	Loiácono and Margaria (2002)
Paramesius brasiliensis	[parasitoid / host] *Eciton burchelli* (Formicidae)	Loiácono and Margaria (2002)
Paraspilomicrus	[parasitoid / host] *Lucilia* (Calliphoridae)	Thompson (1955)
Paraspilomicrus froggatti	[gregarious or super / ? - pupal] *Lucilia* (Calliphoridae)	Johnston and Tiegs (1922); Thompson (1955)
	[parasitoid / host] *Lucilia* (Calliphoridae)	Thompson (1955)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Pentapria	[host / parasitoid] Stratiomyidae	Masner and Garcia (2002); Fouts (1939)
Philolestoides wasmanni	[associated with / potential host] *Neivamyrmex legionis* (Formicidae)	Specimen
	[parasitoid / host] *Neivamyrmex legionis* (Formicidae)	Loiácono and Margaria (2002)
Plagiopria passerai	[endoparasitoid / pupal host] *Plagiolepis pygmaea* (Formicidae)	Lachaud and Passera (1982)
	[endoparasitoid / pupal host] *Plagiolepis pygmaea* (Formicidae)	Lachaud and Passera (1982)
Platymischus dilatatus	[parasitoid / host] Coelopidae (Coelopidae)	Baudoin (1949); Baudoin (1952); Masner and Garcia (2002)
	[endoparasitoid / ? - pupal] *Orygma* (Sepsidae)	Specimen
	[parasitoid / host] *Orygma* (Sepsidae)	Masner and Garcia (2002); Nixon (1980); Backlund (1945)
Psilus	[associated with / potential host] Diptera	Specimen
	[endoparasitoid / ? - pupal] *Opius* (Braconidae) [endoparasitoid / larval host] Tephritidae	Clausen et al. (1965)
	[parasitoid / host] *Scathophaga* (Scathophagidae)	Muesebeck (1980)
	[endoparasitoid / ? - pupal] *Bactrocera cucurbitae* (Tephritidae)	Clausen et al. (1965)
	[endoparasitoid / ? - pupal] *Bactrocera dorsalis* (Tephritidae)	Clausen et al. (1965)
	[endoparasitoid / ? - pupal] *Bactrocera jarvisi* (Tephritidae)	Clausen et al. (1965)
	[endoparasitoid / ? - pupal] *Bactrocera tau* (Tephritidae)	Clausen et al. (1965)
	[endoparasitoid / ? - pupal] *Ceratitis capitata* (Tephritidae)	Aliniazee (1975)
	[endoparasitoid / ? - pupal] *Dacus ciliatus* (Tephritidae)	Clausen et al. (1965)
	[endoparasitoid / ? - pupal] *Rhagoletis pomonella* (Tephritidae) [feeding on fruit / host] *Crataegus*	Maier (1981)
	[endoparasitoid / ? - pupal] *Rhagoletis pomonella* (Tephritidae) [feeding on fruit / host] *Pyrus*	Maier (1981)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[gregarious or super / ? - pupal] *Rhagoletis pomonella* (Tephritidae)	Cameron and Morrison (1974)
Rostropria inopicida	[gregarious endoparasitoid / pupal host] *Inopus rubriceps* (Stratiomyidae)	Early and Naumann (1990)
Spilomicrus	[parasitoid / host] *Bibio ferruginatus* (Bibionidae)	Thompson (1955)
	[parasitoid / host] *Bibio hortulanus* (Bibionidae)	Thompson (1955)
	[parasitoid / host] *Taphrorynchus* (Curculionidae)	Thompson (1955)
	[endoparasitoid / ? - pupal] Syrphidae	Masner (1991)
	[parasitoid / host] *Paratheresia claripalpis* (Tachinidae)	Thompson (1955)
Spilomicrus antennatus	[associated with / potential host] *Arion rufus*	Specimen
Spilomicrus barnesi	[associated with / found in soil near] *Neolimnia tranquilla* (Sciomyzidae)	Specimen
	[gregarious or super / ? - pupal] *Neolimnia tranquilla* (Sciomyzidae)	Early and Horning (1978)
Spilomicrus basalyformis	[endoparasitoid / larval host] Staphylinidae	Masner (1991)
Spilomicrus formosus	[endoparasitoid / ? - pupal] *Pipunculus* (Pipunculidae)	Masner (1991); Specimen
Spilomicrus hemipterus	[solitary endoparasitoid / host] *Tephrochlamys tarsalis* (Heleomyzidae)	Munk (1991)
Spilomicrus ikezakii	[endoparasitoid / ? - pupal] *Lathyrophthalmus* (Syrphidae)	Honda (1969)
Spilomicrus inornatus	[endoparasitoid / ? - pupal] Muscidae	Masner (1991)
	[endoparasitoid / ? - pupal] Muscidae	Masner (1991)
Spilomicrus pilgrimi	[parasitoid / ? - pupal] Diptera	Early (1978)
Spilomicrus stigmaticalis	[endoparasitoid / larval host] Staphylinidae	Masner (1991)
Spilomicrus virginicus	[endoparasitoid / larval host] *Xylota bicolor* (Syrphidae)	Masner (1991)
Szelenyiopria lucens	[gregarious endoparasitoid / larval host] *Acromyrmex ambiguus* (Formicidae)	Loiácono (1987)
Szelenyiopria pampeana	[gregarious endoparasitoid / larval host] *Acromyrmex lobicornis* (Formicidae)	Loiácono et al. (2000)
	[gregarious endoparasitoid / larval host] *Acromyrmex lobicornis* (Formicidae)	Loiácono et al. (2000)
Szelenyiopria reichenspergeri	[parasitoid / host] *Eciton quadriglume* (Formicidae)	Loiácono and Margaria (2002)
Szelenyiopria reichenspergeri	[parasitoid / host] *Neivamyrmex legionis* (Formicidae)	Loiácono and Margaria (2002)
Tetramopria aurocincta	[parasitoid / ? - pupal] Diptera [parasitoid / host] Tortricidae	

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
	[parasitoid / ? - pupal] *Hypoderma bovis*	
	[endoparasitoid / pupal host] *Compsilura concinnata* (Tachinidae) [endoparasitoid / pupal host] *Hyphantria cunea*	Szelényi (1957)
	[parasitoid / ? - pupal] Tachinidae	
	[parasitoid / ? - pupal] Tachinidae [parasitoid / host] Chrysomelidae	
	[solitary endoparasitoid / larval host] *Winthemia* (Tachinidae)	
Tetramopria cincticollis	[parasitoid / host] *Winthemia* (Tachinidae)	
Trichopria	[associated with / potential host] Diptera	Specimen
Trichopria	[endoparasitoid / ? - pupal] Diptera	Specimen
Trichopria	[parasitoid / host] *Melanagromyza obtusa* (Agromyzidae) [feeding on fruit / host] *Cajanus cajan*	Thakur and Odak (1982)
Trichopria	[endoparasitoid / ? - pupal] *Delia radicum* (Anthomyiidae)	Specimen
Trichopria	[endoparasitoid / ? - pupal] *Paregle cinerella* (Anthomyiidae)	Figg et al. (1983)
Trichopria	[endoparasitoid / host] *Hylemyia* (Anthomyiidae)	Specimen
Trichopria	[endoparasitoid / host] *Paregle cinerella* (Anthomyiidae)	Figg et al. (1983); Blume (1984)
Trichopria	[endoparasitoid / host] Aphididae	Specimen
Trichopria	[gregarious endoparasitoid / pupal host] *Calliphoridae*	Hoggsette et al. (1994)
Trichopria	[endoparasitoid / larval host] *Orseolia oryzae* (Cecidomyiidae)	Soenarjo (1986)
Trichopria	[parasitoid / host] *Perrisia* (Cecidomyiidae)	Ferriere (1927)
Trichopria	[endoparasitoid / host] *Cephus cinctus* (Cephidae)	
Trichopria	[parasitoid / host] *Oscinella frit* (Chloropidae)	Thompson (1955)
Trichopria	[parasitoid / host] *Diatraea saccharalis* (Crambidae)	Thompson (1955); Jaynes (1933)
Trichopria	[endoparasitoid / ? - pupal] *Drosophila* (Drosophilidae)	Clausen et al. (1965)
Trichopria	[endoparasitoid / pupal host] *Drosophila ananassae* (Drosophilidae)	Kawanishi and Watanabe (1981)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Trichopria	[endoparasitoid / pupal host] *Drosophila auraria* (Drosophilidae)	Kawanishi and Watanabe (1981)
Trichopria	[endoparasitoid / pupal host] *Drosophila melanogaster* (Drosophilidae)	Kawanishi and Watanabe (1981)
Trichopria	[endoparasitoid / pupal host] *Drosophila simulans* (Drosophilidae)	Kawanishi and Watanabe (1981)
Trichopria	[endoparasitoid / pupal host] *Fannia canicularis* (Fanniidae)	Legner (1966); Legner et al. (1967)
Trichopria	[endoparasitoid / pupal host] *Fannia femoralis* (Fanniidae)	Legner (1966); Legner et al. (1967)
Trichopria	[gregarious endoparasitoid / larval host] *Acromyrmex lobicornis* (Formicidae)	Loiácono et al. (2000)
Trichopria	[parasitoid / host] *Glossina pallidipes* (Glossinidae)	Thompson (1955)
Trichopria	[endoparasitoid / host] *Lasiocampidae* (Lasiocampidae)	Specimen
Trichopria	[gregarious or super / ? - pupal] *Mimegralla* (Micropezidae)	Jacob (1980); Ghorpade et al. (1982)
Trichopria	[endoparasitoid / ? - pupal] *Gymnodia* (Muscidae)	Harris and Summerlin (1984)
Trichopria	[endoparasitoid / ? - pupal] *Gymnodia quadristigma* (Muscidae)	Marchiori et al. (2000)
Trichopria	[endoparasitoid / ? - pupal] *Haematobia irritans* (Muscidae)	Combs and Hoelscher (1969); Mackenzie and Richerson (1993)
Trichopria	[endoparasitoid / ? - pupal] *Stomoxys* (Muscidae)	Specimen
Trichopria	[endoparasitoid / host] *Orthellia caesarion* (Muscidae)	Blume (1984); Figg et al. (1983)
Trichopria	[endoparasitoid / pupal host] *Haematobia irritans* (Muscidae)	Mendes and Linhares (1999)
Trichopria	[endoparasitoid / pupal host] *Musca domestica* (Muscidae)	Legner (1966); Legner et al. (1967); Legner and Olton (1968); Legner and Greathead (1969); Skovgard and Jespersen (1999)
Trichopria	[endoparasitoid / pupal host] *Stomoxys calcitrans* (Muscidae)	Legner and Olton (1968); Legner and Greathead (1969); Smith et al. (1987); Smith et al. (1987)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Trichopria	[gregarious endoparasitoid / host] *Stomoxys calcitrans* (Muscidae)	Hoggsette et al. (1994)
Trichopria	[parasitoid / host] *Haematobia exigua* (Muscidae)	Thompson (1955)
Trichopria	[parasitoid / host] *Hydrotaea australis* (Muscidae)	Thompson (1955)
Trichopria	[parasitoid / ? - pupal] Muscidae	
Trichopria	[solitary endoparasite / larval-pupal host] *Atherigona soccata* (Muscidae)	Taley and Thakare (1979)
Trichopria	[solitary endoparasitoid / pupal host] *Haematobia irritans* (Muscidae)	Marchiori (2001)
Trichopria	[solitary endoparasitoid / pupal host] *Stomoxys calcitrans* (Muscidae)	Huggert and Morgan (1993)
Trichopria	[solitary parasitoid / ? - pupal] *Gymnodia arcuata* (Muscidae)	Figg et al. (1983)
Trichopria	[parasitoid / host] *Hypoborus ficus*	Thompson (1955)
Trichopria	[endoparasitoid / ? - pupal] *Psephenus texanus* (Psephenidae)	Brown (1967)
Trichopria	[solitary endoparasite / larval-pupal host] *Psephenus* (Psephenidae)	Brown (1968); Brown (1987)
Trichopria	[solitary endoparasitoid / larval host] *Psephenus texanus* (Psephenidae)	Specimen
Trichopria	[endoparasitoid / ? - pupal] *Pseudosarchophaga affinis* (Sarcophagidae)	Specimen
Trichopria	[endoparasitoid / ? - pupal] *Ravinia* (Sarcophagidae)	Figg et al. (1983)
Trichopria	[endoparasitoid / ? - pupal] *Ravinia derelicta* (Sarcophagidae)	Watts and Combs (1977)
Trichopria	[endoparasitoid / ? - pupal] *Ravinia lherminieri* (Sarcophagidae)	Figg et al. (1983)
Trichopria	[endoparasitoid / ? - pupal] *Ravinia querula* (Sarcophagidae)	Figg et al. (1983)
Trichopria	[endoparasitoid / ? - pupal] *Sarcophagula occidua* (Sarcophagidae)	Marchiori et al. (2000)
Trichopria	[endoparasitoid / ? - pupal] *Palaeosepsis* (Sepsidae)	Marchiori et al. (2000)
Trichopria	[endoparasitoid / ? - pupal] *Sepsis biflexuosa* (Sepsidae)	Figg et al. (1983)
Trichopria	[endoparasitoid / host] *Sepsis neocynipsea* (Sepsidae)	Blume (1984)
Trichopria	[parasitoid / host] *Saltella sphondylii* (Sepsidae)	Figg et al. (1983)
Trichopria	[endoparasitoid / ? - pupal] Sphaeroceridae	Marchiori et al. (2000)
Trichopria	[associated with / potential host] *Euparyphus (Nigriparyphus) ornatus* (Stratiomyidae)	Specimen
Trichopria	[endoparasitoid / ? - pupal] Stratiomyidae	Bradley et al. (1984)
Trichopria	[endoparasitoid / host] *Hermetia illucens* (Stratiomyidae)	Specimen
Trichopria	[endoparasitoid / larval host] *Hermetia illucens* (Stratiomyidae)	Bradley et al. (1984)
Trichopria	[endoparasitoid / pupal host] *Hermetia illucens* (Stratiomyidae)	Bradley et al. (1984)
Trichopria	[endoparasitoid / pupal host] *Hermetia illucens* (Stratiomyidae)	Mitchell et al. (1974)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Trichopria	[gregarious endoparasitoid / pupal host] *Microdon albicomatus* (Syrphidae) [predator / prey] *Camponotus noveboracensis* (Formicidae)	Paulson and Akre (1991)
Trichopria	[endoparasitoid / ? - pupal] Tabanidae	
Trichopria	[endoparasitoid / host] *Tabanus reinwardtii* (Tabanidae)	Specimen
Trichopria	[associated with / potential host] Tachinidae	Specimen
Trichopria	[endoparasitoid / ? - pupal] *Bessa selecta* (Tachinidae)	Specimen
Trichopria	[endoparasitoid / host] *Myxexoristops hertingi* (Tachinidae)	
Trichopria	[endoparasitoid / host] *Myxexoristops hertingi* (Tachinidae) [endoparasitoid / host] *Acantholyda erythrocephala* (Pamphiliidae)	Specimen
Trichopria	[endoparasitoid / host] Tachinidae	Specimen
Trichopria	[endoparasitoid / host] Tachinidae [parasitoid / host] Coleoptera	Specimen
Trichopria	[endoparasitoid / pupal host] Tachinidae [endoparasitoid / host] *Bombyx mori* (Bombycidae)	Veeranna et al. (1987)
Trichopria	[gregarious or super / ? - pupal] *Doleschalla* (Tachinidae) [endoparasitoid / larval host] *Pantorhytes szentivanyi*	Baker (1978)
Trichopria	[parasitoid / host] *Lixophaga diatraeae* (Tachinidae)	Myers (1931); Thompson (1955)
Trichopria	[parasitoid / host] *Palpozenillia palpalis* (Tachinidae)	Myers (1935); Thompson (1955)
Trichopria	[endoparasitoid / host] *Pristiphora erichsonii* (Tenthredinidae)	Specimen
Trichopria	[endoparasitoid / ? - pupal] *Anastrepha suspensa* (Tephritidae)	Baranowski et al. (1993)
Trichopria	[endoparasitoid / ? - pupal] *Bactrocera cucurbitae* (Tephritidae)	Clausen et al. (1965)
Trichopria	[endoparasitoid / ? - pupal] *Bactrocera dorsalis* (Tephritidae)	Clausen et al. (1965)
Trichopria	[endoparasitoid / ? - pupal] *Bactrocera incisus* (Tephritidae)	Clausen et al. (1965)
Trichopria	[endoparasitoid / ? - pupal] *Bactrocera tau* (Tephritidae)	Clausen et al. (1965)
Trichopria	[endoparasitoid / ? - pupal] *Ceratitis anonae* (Tephritidae)	Clausen et al. (1965)
Trichopria	[endoparasitoid / ? - pupal] *Dacus ciliatus* (Tephritidae)	Clausen et al. (1965)
Trichopria	[endoparasitoid / ? - pupal] Tephritidae	Clausen et al. (1965)
Trichopria	[endoparasitoid / pupal host] *Anastrepha* (Tephritidae)	Ovruski et al. (2000)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Trichopria	[endoparasitoid / pupal host] *Bactrocera dorsalis* (Tephritidae)	Puttarudriah and Usman (1954)
Trichopria	[parasitoid / host] *Anastrepha* (Tephritidae)	Narayanan and Chawla (1962)
Trichopria	[parasitoid / host] *Anastrepha fraterculus* Wiedemann (Tephritidae) [feeding on fruit / host] *Citrus sinensis*	Raga et al. (2004)
Trichopria	[parasitoid / host] *Anastrepha obliqua* (Tephritidae) [on / host] *M. indica*	Jiron and Mexzon (1989)
Trichopria	[parasitoid / host] *Anastrepha striata* (Tephritidae) [on / host] *P. guajava*	Jiron and Mexzon (1989)
Trichopria	[parasitoid / host] *Bactrocera incisus* (Tephritidae)	Puttarudriah and Usman (1954)
Trichopria aequata	[parasitoid / host] *Oscinella frit* (Chloropidae)	Thompson (1955)
Trichopria aequata	[endoparasitoid / pupal host] *Drosophila kuntzei* (Drosophilidae) [feeding on fruit / host] *Allium ursinum*	Offenberger and Klarenberg (1994)
Trichopria aequata	[endoparasitoid / pupal host] *Drosophila limbata* (Drosophilidae) [feeding on fruit / host] *Allium ursinum*	Offenberger and Klarenberg (1994)
Trichopria aequata	[endoparasitoid / pupal host] *Drosophila phalerata* (Drosophilidae) [feeding on fruit / host] *Allium ursinum*	Offenberger and Klarenberg (1994)
Trichopria aequata	[endoparasitoid / pupal host] *Drosophila transversa* (Drosophilidae) [feeding on fruit / host] *Allium ursinum*	Offenberger and Klarenberg (1994)
Trichopria aequata	[associated with / potential host] *Formica* (Formicidae)	
Trichopria aequata	[solitary endoparasitoid / ? - pupal] *Megaselia agarici* (Phoridae) [feeding on fruit / host] *Agaricus augustus*	Notton (1991)
Trichopria anastrephae	[endoparasitoid / larval host] *Anastrepha fraterculus* Wiedemann (Tephritidae) [feeding on fruit / host] *P. guajava*	García and Corseuil (2004)
Trichopria anastrephae	[endoparasitoid / larval host] *Anastrepha fraterculus* Wiedemann (Tephritidae) [feeding on fruit / host] *Spondias mombin*	Aguiar-Menezes et al. (2001)
Trichopria anastrephae	[parasitoid / host] *Anastrepha* (Tephritidae)	Ovruski et al. (2000); Loiácono and Margaria (2002)
Trichopria anastrephae	[parasitoid / host] *Anastrepha* (Tephritidae) [on / host] *Anastrepha serpentina* (Tephritidae)	Narayanan and Chawla (1962)
Trichopria anastrephae	[parasitoid / host] *Anastrepha* (Tephritidae) [on / host] *Spondias dulcis*	Narayanan and Chawla (1962)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Trichopria anastrephae	[parasitoid / host] *Anastrepha serpentina* (Tephritidae)	Loiácono and Margaria (2002)
Trichopria angustipennis	[endoparasitoid / pupal host] *Lemnaphila scotlandae* (Ephydridae) [feeding on leaves / host] *Lemna valdiviana*	Buckingham (1989)
Trichopria atrata	[solitary endoparasitoid / pupal host] Sphaeroceridae	Notton (1994)
Trichopria atrichomelinae	[gregarious endoparasitoid / pupal host] *Atrichomelina pubera* (Sciomyzidae)	O'Neill (1973)
Trichopria brevipennis	[endoparasitoid / ? - pupal] *Pollenia rudis* (Calliphoridae)	Kieffer (1911)
Trichopria capensis	[endoparasitoid / pupal host] *Ceratitis capitata* (Tephritidae)	Silvestri (1914)
Trichopria capensis	[parasitoid / host] *Ceratitis* (Tephritidae)	Narayanan and Chawla (1962)
Trichopria catarinensis	[parasitoid / host] *Eciton burchelli* (Formicidae)	Loiácono and Margaria (2002)
Trichopria cilipes	[parasitoid / host] *Agromyza potentillae* (Agromyzidae)	Thompson (1955)
Trichopria cilipes	[gregarious or super / ? - pupal] *Calliphora* (Calliphoridae)	Teodorescu and Ursu (1979)
Trichopria cilipes	[gregarious or super / ? - pupal] *Calliphora erythrocephala* (Calliphoridae)	Teodorescu and Ursu (1979)
Trichopria cilipes	[gregarious or super / ? - pupal] *Lucilia* (Calliphoridae)	Teodorescu and Ursu (1979)
Trichopria cilipes	[gregarious or super / ? - pupal] *Piophila casei* (Piophilidae)	Teodorescu and Ursu (1979)
Trichopria columbiana	[solitary endoparasitoid / pupal host] *Hydrellia balciunasi* (Ephydridae) [feeding on leaves / host] *Hydrilla verticillata*	Coon (2000)
Trichopria columbiana	[solitary endoparasitoid / pupal host] *Hydrellia pakistanae* (Ephydridae) [feeding on leaves / host] *Hydrilla verticillata*	Coon (2000)
Trichopria commoda	[endoparasitoid / pupal host] *Aldrichina grahami* (Calliphoridae)	Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Lucilia* (Calliphoridae)	Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Lucilia cuprina* (Calliphoridae)	
Trichopria commoda	[endoparasitoid / pupal host] *Lucilia cuprina* (Calliphoridae)	Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Lucilia illustris* (Calliphoridae)	Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Musca domestica* (Muscidae)	Muesebeck (1961); Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Muscina stabulans* (Muscidae)	Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Bellieria melanura* (Sarcophagidae)	Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Liopygia crassipalpis* (Sarcophagidae)	Yasumatsu (1964)
Trichopria commoda	[endoparasitoid / pupal host] *Sarcophaga peregrina* (Sarcophagidae)	Yasumatsu (1964)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Trichopria commoda	[endoparasitoid / pupal host] *Eristalis cerealis* (Syrphidae)	Yasumatsu (1964)
Trichopria crassifemur	[parasitoid / ? - pupal] *Hylemyia* (Anthomyiidae)	
Trichopria cubensis	[parasitoid / host] *Lixophaga diatraeae* (Tachinidae)	Thompson (1955)
Trichopria cubensis	[parasitoid / host] *Lydella* (Tachinidae)	Loiácono and Margaria (2002)
Trichopria cubensis	[parasitoid / host] *Paratheresia brasiliensis* (Tachinidae)	Loiácono and Margaria (2002)
Trichopria cubensis	[parasitoid / host] *Parthenoleskia parkeri* (Tachinidae)	Loiácono and Margaria (2002)
Trichopria drosophilae	[endoparasitoid / pupal host] *Drosophila melanogaster* (Drosophilidae)	Romani et al. (2002)
Trichopria formicaria	[parasitoid / host] *Formica fusca* (Formicidae)	Kieffer (1911)
Trichopria grenadensis	[parasitoid / host] *Hermetia illucens* (Stratiomyidae)	Loiácono and Margaria (2002)
Trichopria hirticollis	[parasitoid / host] Calliphoridae	Thompson (1955)
Trichopria hyalinipennis	[solitary endoparasitoid / host] *Tephrochlamys tarsalis* (Heleomyzidae)	Munk (1991)
Trichopria hyalinipennis	[endoparasitoid / pupal host] *Compsilura concinnata* (Tachinidae) [endoparasitoid / pupal host] *Hyphantria cunea*	Szelényi (1957)
Trichopria lamellifera	[parasitoid / host] Diptera	Loiácono and Margaria (2002)
Trichopria lewisi	[endoparasitoid / ? - pupal] *Glossina* (Glossinidae)	Nixon (1940)
Trichopria lewisi	[endoparasitoid / ? - pupal] *Glossina brevipalpis* (Glossinidae)	Nixon (1940)
Trichopria lonchaearum	[gregarious or super / ? - pupal] *Piophila casei* (Piophilidae)	Teodorescu and Ursu (1979)
Trichopria major	[solitary parasitoid / ? - pupal] *Paregle* (Anthomyiidae)	Teodorescu and Ursu (1979)
Trichopria major	[gregarious or super / ? - pupal] *Lucilia* (Calliphoridae)	Teodorescu and Ursu (1979)
Trichopria major	[solitary parasitoid / ? - pupal] *Piophila casei* (Piophilidae)	Teodorescu and Ursu (1979)
Trichopria myrmicae	[associated with / potential host] *Myrmica americana* (Formicidae)	Specimen
Trichopria nigra	[solitary parasitoid / ? - pupal] *Piophila casei* (Piophilidae)	Teodorescu and Ursu (1979)
Trichopria paludis	[endoparasitoid / pupal host] *Lemnaphila scotlandae* (Ephydridae) [feeding on leaves / host] *Lemna valdiviana*	Buckingham (1989)
Trichopria popei	[gregarious endoparasitoid / pupal host] *Atrichomelina pubera* (Sciomyzidae)	O'Neill (1973)
Trichopria popei	[gregarious or super / ? - pupal] *Elgiva solicita* (Sciomyzidae)	Knutson and Berg (1963)
Trichopria popei	[gregarious or super / ? - pupal] *Sepedon* (Sciomyzidae)	Muesebeck (1949)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Trichopria popei	[super endoparasitoid / pupal host] *Dictya* (Sciomyzidae)	Knutson and Berg (1963)
Trichopria rotundata	[endoparasitoid / pupal host] *Compsilura concinnata* (Tachinidae) [endoparasitoid / pupal host] *Athalia rosae* (Tenthredinidae)	Szelényi (1957)
Trichopria rotundata	[endoparasitoid / pupal host] *Compsilura concinnata* (Tachinidae) [endoparasitoid / pupal host] *Hyphantria cunea*	Szelényi (1957)
Trichopria rubrithoraca	[found with / associated with] Muridae	
Trichopria stomoxydis	[gregarious endoparasitoid / pupal host] *Stomoxys calcitrans* (Muscidae)	Hogsette et al. (1990); Hogsette et al. (1990); Vaughan (1985)
Trichopria subimpressa	[endoparasitoid / ? - pupal] *Microchrysa polita* (Stratiomyidae)	Notton (1991)
Trichopria subimpressa	[endoparasitoid / ? - pupal] *Syritta pipiens* (Syrphidae)	Notton (1991)
Trichopria subpetiolata	[endoparasitoid / ? - pupal] *Eristalis* (Syrphidae)	Honda (1969)
Trichopria tabanivora	[endoparasitoid / pupal host] *Chrysops* (Tabanidae)	Segal (1936)
Trichopria tabanivora	[gregarious or super / ? - pupal] *Chrysops mitis* (Tabanidae)	Cameron (1926); Jones and Anthony (1964)
Trichopria tabanivora	[gregarious or super / ? - pupal] *Tabanus nigrovittatus* (Tabanidae)	Bailey (1947)
Trichopria tabanivora	[gregarious or super / ? - pupal] *Tabanus reinwardtii* (Tabanidae)	Cameron (1926); Jones and Anthony (1964)
Trichopria tetratoma	[gregarious or super / ? - pupal] *Piophila casei* (Piophilidae)	Teodorescu and Ursu (1979)
Trichopria verticillata	[endoparasitoid / ? - pupal] *Chloropidae* (Chloropidae)	Notton (1991)
Trichopria verticillata	[found with / associated with] *Formica* (Formicidae)	
Ismarinae		
Ismarus dorsiger	[endoparasitoid / ? - pupal] *Aphelopus* (Dryinidae) [parasitoid / host] *Fagocyba cruenta*	Jervis (1979)
Ismarus dorsiger	[endoparasitoid / ? - pupal] *Aphelopus* (Dryinidae) [parasitoid / host] *Ribautiana ulmi*	Jervis (1979)
Ismarus flavicornis	[solitary endoparasitoid / host] *Anteon flavicorne* (Dryinidae) [solitary ectoparasitoid / host] *Idiocercus*	Chambers (1955)

Table 4.3 Continued.

Parasitoid	[Relationship] / Host(s)	Citation
Ismarus halidayi	[solitary endoparasitoid / host] *Anteon* (Dryinidae) [solitary ectoparasitoid / host] *Oncopsis*	Chambers (1955)
Ismarus halidayi	[solitary endoparasitoid / host] *Anteon infectum* (Dryinidae) [solitary ectoparasitoid / host] *Iassus lanio* [on / host] *Quercus robur*	Chambers (1981)
Ismarus neotropicus	[parasitoid / host] Dryinidae	Loiácono and Margaria (2002)
Ismarus rugulosus	[endoparasitoid / ? - pupal] *Chelogynus* (Dryinidae) [parasitoid / host] *Streptanus sordidus*	Jervis (1979)
Ismarus varicornis	[parasitoid / host] Dryinidae	Loiácono and Margaria (2002)

While the general mode and scope of parasitism is thus largely that recognized by Masner (1993), accumulation of specific host records has revealed that diapriids have perhaps a more diverse host range than previously expected. There are confirmed host records for over 35 families of Diptera, two beetle families (Psephenidae, Staphylinidae) and two hymenopteran families (Dryinidae, Formicidae).

Several records of gregarious parasitism exist for the Diapriinae (e.g. *Trichopria*, *Basalys*, *Spilomicrus*- and several purportedly related genera, see Table 4.3). The gregarious Diapriinae are all purportedly highly derived taxa. No records of gregarious parasitism are presently known for the Belytinae. No records of egg parasitism have been recorded for the family.

Over 30 records of ant-diapriid associations are presently known. Of these records many do not mention specific observations of ant-parasitism but rather simply ant-diapriid associations, (e.g. found in columns of army ants, Huggert and Masner (1983). Masner and Garcia (2002) provide summary information on the majority of these types of relationships. In particular, while Loiácono and Margaria (2002) list as hosts for diapriids many ant records there is no mention as to the distinction between parasitism and simple association. Conclusive evidence of diapriid-ant parasitism is available in Loiácono (1987), Loiácono et al. (2000) and Fernández-Marín et al. (2006).

Unfortunately no published phylogeny exists for the Diapriidae and the phylogeny presented in Chapter VI is based on an inadequate sampling of taxa with known biologies. Hypotheses of biological evolution are therefore largely subjective. We can however make several observations, noting that these require a well resolved phylogeny to test further. It is presently unclear as to the precise ground-plan biology of the family. This is largely due to two factors, a lack of records for purportedly basal members of the family, and a lack of carefully documented data for those records that exist. With respect to the latter, to reiterate, many diapriid host records have been gathered indirectly, and whether the host remains have been properly isolated and identified in these cases is unclear. For example, at least some diapriid species are hyperparasitoids, *Trichopria* in particular (Baker, 1978; Ghorpade et al., 1980; Paulson

and Akre, 1991; Hoggsette et al., 1994). Knowing this, many families with known diapriid associations may not be primary hosts, i.e. all recorded lepidopteran families, the beetle family Curculionidae, an aphid, and hymenopteran families Eulophidae, Tenthredinidae, Pamphiliidae and Cephidae. Furthermore, the distinction between an endo- and ectoparasitoid is non-trivial in the case of pupal parasitism, as a parasitoid can attach external to the host but still be inside the pupal exuviae (in the case of Nematocera) or puparium (in the case of cyclorrhaphous flies). Given these factors, while most records suggest a ground plan biology of solitary pupal endoparasitism, this is likely only due to the fact that emergence frequently occurs at the pupal stage, i.e. the exact stage and mode of parasitism has rarely been recorded. Careful observations (e.g. Hellqvist, 1994) have shown that at least some diapriids parasitize the larval stage.

Though found in many agricultural settings diapriids have not traditionally been considered as potentially useful biological control agents (but see Vaughan, 1985; Notton, 1997). In at least several cases they are known to interfere with potential controlling agents through hyperparasitism (e.g. Wright et al., 1947; Veeranna et al., 1987). In most cases the number of recovered individuals in these studies hints that diapriids will not disrupt biological control efforts in any important manner.

While not generally considered for use in biological control, diapriids are frequently encountered during surveys for potential control agents. For example they have been found in surveys for parasitoids of fruit flies (Silvestrii, 1914; Maier, 1981; Jiron and Mexzon, 1989; Lopez et al., 1999; Sivinski et al., 2000; Wharton et al., 2000), horn fly (Combs and Hoelscher, 1969; Thomas and Morgan, 1972; Harris and Summerlin, 1984; Mackenzie and Richerson, 1993; Mendes and Linhares, 1999), sugar cane borer (Jaynes, 1933), carrot fly (Wright et al., 1947), filth flies (Legner, 1966, Legner and Olton, 1968, 1971; Legner and Greathead, 1969; Mitchel et al., 1974; Figg et al., 1983; Blume, 1984; Hoggsette et al., 1994; Floate et al., 1999; Skovgard and Jespersen, 1999) and in broader geographically based surveys (e.g. Myers, 1935). While the number of recovered diapriid individuals in these studies are nearly universally very low (typically less than 5%), their wide-spread presence is notable.

Several diapriid species have been released for control, these belonging primarily to the genera *Coptera* and *Trichopria* (see Clausen et al. 1965), also notably against the greenhouse pest fly *Bradesia* (Notton 1997).

The present project's primary goal is to record the known host records for the Diapriidae. Future expansions of the project are planned and these will progress towards several goals. While diapriids may ultimately have only minimal application as biological control agents their nearly ubiquitous distribution suggests they may be useful as indicators of environmental health. In certain habitats, especially within forests, diapriids may be the most numerically abundant parasitoids. Given this, future efforts will revolve around identifying literature which may help to identify habitats whose health can be monitored by the diapriid fauna therein. This will be done using the keyword indexing functionality presently available in the database. Ultimately, we plan to use mx's capability to capture specimen level data including collecting event data to provide and estimate host and parasitoid distributions. Finally, we hope to become a catalog of all biological information (e.g. phenology, functional morphology, behaviour) not just that which is host related. In this regard the hypotheses on the evolution of myrmecophily proposed by Huggert and Masner (1983) are particularly interesting. A core database of observations on myrmecophily will facilitate the testing of these hypotheses.

Smaller, less well known groups of organisms traditionally have only sporadic work done on them, with the time between subsequent efforts frequently extensive. This is reflected by the large gaps in host record catalogs for the Diapriidae (see chronological listing of associations online at http://www.diapriid.org/projects/4/public/association/browse_chronological). Databasing efforts can bridge these periods of inactivity. Problems with lag between efforts are further exacerbated by the relative paucity of people studying the biology of these groups. Unified efforts and repositories are essential to increasing our understanding of these groups. These efforts allow new workers to rapidly review past work and identify areas wherein new effort will be maximally useful. For example,

given the diapriid data presented here it is clear that perhaps the most important data to seek out pertains to the hosts of the Ambositrinae, of which only a single over 100 year old record is known. Databases further allow for mistakes or misinterpretations to be rapidly identified and clarified, a process which is ongoing for the present work. The database application provided here seeks to address these issues, it allows for a truly global collaborative effort, with experts being able to add and review the underlying data over the world-wide web. As diapriid information is rarely encountered and likely frequently lost or overlooked we make a plea for unification of these data and invite potential collaborators to join the project via direct or indirect submission of data, or by providing critical feedback.

Acknowledgements

This work was supported by NSF PEET grant DEB-0328922 and REU grant DEB-0522836 both to Bob Wharton. Jonathon Cammack and Krishna Dole helped in the preparation of this paper. Several people helped by providing host records, feedback on the application, or data entry: Andy Deans, Katja Seltmann, Jacques Dubois, David Notton, Victor Kolyada, Lubomir Masner, Marta Loiácono, and Terry Junek. Thanks to Jim Woolley, Bob Wharton and Amy Bader for valuable feedback that greatly improved this paper.

CHAPTER V

HYPER-PARTITIONING, PAIRING, AND PARSIMONY- ANALYSIS OF rRNA DATA WITH AN EMPIRICAL EXAMPLE FROM THE DIAPRIIDAE (HYMENOPTERA)

Overview

Ribosomal RNA (rRNA) sequences, popular for use in phylogenetic analyses, typically possess both a large number of indels and numerous covarying (evolutionary non-independent) positions. Indels are a multifaceted problem for phylogenetic analyses, as they potentially represent both phylogenetic signal and high levels of homoplasy. We provide a simple alignment format that addresses some of these issues (i.e. incorporates hypotheses of structure) and a set of scripts that facilitates the analysis of data in this format. Data in this format is hyper-partitioned, that is, it may contain hundreds of partitions delineated by structural criteria. The evolutionary non-independence of basepairs within rRNA molecules is another confounding problem, and is oftentimes overlooked in phylogenetic inference. This non-independence violates various statistical assumptions and philosophical underpinnings of phylogenetic reconstruction. We show that a majority of sites in the nuclear large (28S) and small (18S) rRNA subunits, extremely popular for phylogenetic analyses, are demonstrably non-independent. The effect of partitioning, and the further incorporation of hypotheses of basepairing is explored under the parsimony criterion. We provide two simple parsimony methods that are minor variants of the ARC scripts (Kauff et al. 2003), and a third that is based on the 20 state model of Smith et al. (2004). All three are tested on a dataset of 28S (expansion segments D2-D5) and 18S (variable region V4) rRNA for the Diapriidae, a family of wasps for which no quantitative phylogeny exists. The results of our analyses show that: 1) partition-level homology statements corresponding to the secondary structure of the molecule are capable of recovering relationships recovered using columnar homology statements (for the same data); 2) accounting for the non-independence of sites decreases clade support; and 3) the partition and alignment-free recoding methods implemented

here provide results highly congruent with those under standard columnar-alignment methods. The recovered suprageneric relationships among diapriids broadly concur with observations from morphological data; however, various novel relationships are recovered and discussed.

Introduction

Ribosomal RNA (rRNA) is widely used in phylogenetic inference (Adoutte et al., 2000) because of its ubiquity across life and its relative ease of extraction and amplification due to high copy number (Noller and Woese, 1981). Ribosomal RNA molecules contain conserved regions that evolve at slower rates and variable regions that evolve at faster rates (van de Peer et al., 1993). The presence of rate-diversity in a single molecule, which translates to signal at different depths of a phylogeny, is a desirable characteristic of a molecular marker. Extracting the signal from these molecules remains problematic (Morrison, 2006). While conserved regions (e.g. helices, or those sites typically involved in base pairing) are relatively easily homologized there is much debate as to how variable length regions (e.g. loops) should be identified and integrated into phylogenetic analyses (Lee, 2001; Smythe et al., 2006; and see review in Morrison, 2006).

Methods to identify and delimit regions of high levels of indels are also varied, though few approaches are widely used throughout phylogenetic analysis. Most approaches, whether algorithmic or "manual", rely on the method of identifying conserved motifs, using these as boundaries with which to partition the dataset for further analysis. In cases of rRNA alignments these conserved motifs typically represent known secondary structures, though this fact is not always recognized when alignment protocols are justified. While conserved motifs are typically identifiable when compared with known structures, for example as cataloged at the Comparative RNA Web site (Cannone et. al., 2002), there is generally little effort in phylogenetic inference to formally identify these regions (sensu Kjer, 1995; Gillespie, 2004) and even more rarely are those more variable structures internal to these conserved motifs annotated.

The analysis of rRNA data is further complicated by the high degree of non-independence of sites (Table 5.1, compare mask pairs times 2 vs. aligned columns). The importance of using only independently evolving characters in phylogenetic inference is well documented (Wheeler and Honeycutt, 1988; and see Phillips, 2005), whereas the consequences of using non-independent data in phylogenetic inference are less understood. While in many cases it is difficult to know or test whether characters are in fact non-independent (Phillips, 2005), this is not the case for rRNA molecules whose secondary structure involves pairing between two sites. In these molecules it is not uncommon to find upwards of 60% of the sites used in phylogenetic inference involved in pairing (Table 5.1). This extremely high degree of non-independence is typically not accounted for in parsimony-based analyses (including direct-optimization or fixed-states optimization under parsimony) and can be responsible for variation in support values (Cummings et al., 1995).

Alignment tools that incorporate structural elements (e.g. thermodynamic or base-pairing information) remain a minority as compared to those that don't (Edgar and Batzoglou, 2006; Kjer et al. 2007). There are, however, approaches to multiple sequence alignment that incorporate secondary structure (e.g., Page, 2000; Gardener and Giegerich, 2004; Dowell and Eddy, 2006; and see reviews in Morrison 2006). With respect to the algorithms involved, both alignment and simultaneous prediction of structure (e.g. manipulating hypotheses of basepairing or thermodynamic information) are presently highly technical, and require intensive computation capabilities (Gardener and Giegerich, 2004). Many of these methods also require carefully curated alignments to train the software. Unfortunately, many promising algorithmic approaches are computationally restricted in either the number of taxa (frequently 2, e.g. RAGA, Notredame et. al., 1997 and Dynalign, Mathews and Turner, 2000) or nucleotides that can be used (Gardener and Giegerich, 2004). These limits constrain potential use of structure-based alignment algorithms in modern problems of phylogenetic inference.

TABLE 5.1. A summary of datasets that use the alignment format presented herein. Many of the datasets, including all those published in the table, are available at http://hymenoptera.tamu.edu/rna. Partitions are given as: Total of all (number of total bracketed). Data in bracketed partitions is unaligned, i.e. not aligned manually.

Taxon	Gene(s)	Taxa	Partitions (bracketed)	Columns (aligned/unaligned)	Helices	Mask pairs	Source
Arthropoda	18S	175	342 (84)	5019 (1819/3170)	148	549	Gillespie et al. (2005a)
Hymenoptera	28S D2-D3	84	103 (29)	487 (307/487)	25	146	*MJY and JJG expanding Dowton and Austin (2001)
Hymenoptera	18S	286	292 (241)	2128 (1887/241)	72	556	*MJY and JJG
Ichneumonoidea (Hymenoptera)	28S D2-D10; 18S	290	665 (100)	4912	149	1025	Gillespie et al. (2005b)
Exodontiella (Hymenoptera: Braconidae)	28S	59	112 (9)	566 (517/49)	27	144	Wharton et al. (2006)
Chalcidoidea (Hymenoptera)	28S D2-D3	527	226 (59)	1417 (925/492)	50	282	Gillespie et al. (2005c)
Azotinae (Hymenoptera: Chalcidoidea)	28S D2-D3; 18S	174	273 (66)	1778 (1129/649)	120	331	James Munro (University of California at Riverside, Riverside, California, pers. comm.)
Evaniidae (Hymenoptera: Evanioidea)	28S D2, 16S	60	212 (149)	2648 (2000/648)	45	250	Deans et al. (2006)
Encyrtidae (Hymenoptera: Chalcidoidea)	28S D2-D3	116	138 (44)	985 (613/372)	64	208	* Noyes et al. (with MJY and JJG)
Diapriidae (Hymenoptera: Proctotrupoidea)	28S, 18S	168	987 (54)	7089 (6111/978)	249	1705	Present work.
Culicidae	16S	11	73 (0)	(429/0)	12	88	*JJG.
Scolytidae (Coleoptera)	28S D2-D5	109	219 (41)	1630 (1052/578)	50	340	*Anthony Cognato (Michigan State University, East Lansing, Michigan) and JJG.

* These datasets are presently on-line, but have not been published.

Table 5.1 Continued.

Taxon	Gene(s)	Taxa	Partitions (bracketed)	Columns	Helices	Mask pairs	Source
Galerucinae (Coleoptera: Chrysomelidae)	28S D2-D3	231	146 (29)	854 (683/171)	33	201	Gillespie et al. (2004b)
Galerucinae - Alignment 2 (Coleoptera: Chrysomelidae)	28S D2-D3	249	322 (41)	2648 (2143/505)	72	467	*JJG
Deuterostomes	28S, 18S, 5.8S	48	1002 (191)	7951 (4644/3307)	236	1338	Jon Mallat (Washington State University, Pullman, Washington) and MJY
Vespoidea (Hymenoptera)	18S, 28S	30	284	1790 (1607/183)	70	470	Hines et. al. (2007)
Scleractinia, Actiniaria and Corallimorpharia (Cnidaria: Hexacorallia)	28S	138	166 (32)	967 (696/271)	72	175	Marcos Barbeitos (SUNY at Buffalo, Buffalo, New York)
Scleractinia, Actiniaria and Corallimorpharia (Cnidaria: Hexacorallia)	12S	141	184 (45)	991 (586/405)	70	148	Marcos Barbeitos (SUNY at Buffalo, Buffalo, New York)
Mites (Acarina)	28S, 18S	62	447 (3)	3823 (3809/14)	106	783	Ashley Dowling, (University of Kentucky, Lexington, Kentucky, pers. comm.)
HymATOL Hymenoptera	28S, 18S	97	836 (34)	6322 (5282/1040)	209	1443	Ashley Dowling, (University of Kentucky, Lexington, Kentucky, pers. comm.)

* These datasets are presently stored at the aforementioned web-site, but have not been published.

One solution to the alignment problem has been, and continues to be, to align the data "by-brain" (we prefer to differentiate from the commonly used "by-eye"). This approach remains controversial for various reasons (see following criticisms and those of Ogden et al., 2005). A common argument, that the manual process is subjective and unrepeatable is easily debunked (Kjer et al., 2007) by noting that anyone can take an end-product structural alignment and test or overturn the hypotheses therein. Furthermore, manual alignment techniques are responsible for various very large multiple sequence alignments (Brown, 1999; Wuyts et al. 2001, 2002; Cannone et al., 2002; Griffiths-Jones et al., 2005) that are critically important to benchmarking and methodological explorations (e.g. Gardner et al., 2004, 2005). To suggest that these benchmarks are meaningless because of their manual derivation is an oversimplification. Finally, the manual hypotheses of structural multiple sequence alignments should be an iterative process with various stages of development (Woese et al., 1983). This is the classical framework for developing strong hypotheses of homology, the checking and rechecking proposed by Hennig (1966). Each sequence added to a structure-based alignment provides a test of the current structural hypotheses. In contrast, it is unclear that the process of using algorithmic alignment only inherently includes a stage of hypothesis checking and potential strengthening. Adding more sequences to an algorithm-based alignment does concur with Hennig's proposal to add data to check homology, however, it is unclear as to whether 1) the same (columnar) homology statements are actually present with each addition of a new sequence and therefor actually tested or 2) in the case of dynamic-homology, where the partition is the primary homology statement, what can be learnt about the homology statement itself following the addition of new data. Other criticisms, particularly those relating to the feasibility of applying manual methods to the logarithmically growing quantities of data (Ogden et al., 2005), are well founded, and of greater concern. Structural alignments identify with very high accuracy (> 90%, Gutell et al., 2002) those regions known to be base-paired in the atomic crystal structures (Cate et al., 1999; Ban et al., 2000; Schluenzen et al., 2000; Wimberly et al., 2000; Spahn et al., 2001; Yusupov et al., 2001). These regions are those that would generally be identified by algorithmic means (for a practical example see Hickson et al. 2000). This level of

accuracy competes very well with the algorithms benchmarked by Gardener and Giegerich (2004). Note that the benchmarks therein treated very small structures (e.g. 21 basepairs), whereas there can frequently be an order of magnitude more pairs in even small sized alignments (Table 5.1).

As evidenced in several recent papers (Deans et al. 2006; Gillespie et al. 2005abc, 2006; Wharton et al. 2006) the end product of the manual process can be a highly annotated alignment. As these alignments contain a large number (Table 5.1) of smaller partitions, we term them "hyper-partitioned" and here review the basis for their construction. The number of partitions can be an order of magnitude greater than typically employed for highly partitioned analyses performed in approaches that benefit from highly partitioned data (e.g. POY, Wheeler et. al., 2003). For example, Terry and Whiting (2005) use 26 and 28 partitions and Ogden and Whiting (2005) use 7 and 10 partitions for their respective 18S and 28S rRNA data, whereas comparatively sized amplicons in Table 5.1 may be have over 200 partitions. Hyper-partitioned datasets introduce another factor to issues pertaining to partition choice and construction and congruence among partitions (e.g. ILD metrics), long-standing issues in phylogenetic inference. Partitions in molecular analyses typically correspond to the bounding primers, but in some cases (e.g. rRNAs) there is little a-priori reason to suspect that rates within a given amplicon are uniform. That is, equating the bounds of a partition with the bounds of an amplicon is not necessarily a reasonable assumption. A second important property of hyper-partitioned alignments is that they can mix aligned and unaligned data, for example portions of the molecule which are not involved in base-pairing are explicitly annotated as such (see "unalignability") below.

Methods available for analyzing partitioned alignments have focused on recoding ambiguously aligned regions therein, with data subsequently analyzed under parsimony (fixed-states optimization, Wheeler, 1999; INNASE, Lutzoni et al., 2000). Likelihood and Bayesian (e.g. PHASE, Jow et al. 2002; Hudelot et al. 2003) methods are also available to model hypotheses of base-pairing that may also be present in these alignments.

This paper takes an empirical approach to exploring issues of alignment, character recoding, partitioning, and independence of data under the parsimony criterion. We use

as a case study 28S and 18S rRNA sequence data for taxa of the family Diapriidae (Hymenoptera) and purported sister groups. Diapriids are small (typically 2-4mm), parasitoid (killing their hosts) wasps that are cosmopolitan in distribution, with approximately 4000 estimated species (Masner, 1993). Only a single published quantitative phylogeny (using morphological characters) exists for a species group of diapriids (Loiácono and Margaria, 2000). There are currently four recognized subfamilies of diapriids (Fig. 5.1) that are relatively well defined morphologically (Masner, 1993, Masner and García, 2002), and there are several well-supported hypotheses of intra-familial clades (Naumann, 1982; Masner and García, 2002). Many of the genera in three of the four subfamilies are also relatively well defined, though taxonomy of the family for non-European species functions largely at the genus-group level.

FIGURE 5.1. Representative diapriids. Clockwise from top left: *Ismarus* sp. (Ismarinae); *Coptera* sp. (Diapriinae: Psilini); *Austroxylabis pictipennis* (Ambositrinae), inset shows small second sternite characteristic of the subfamily; *Masnerosema* (Belytinae); *Spilomicrus* sp. (Dipariinae:Spilomicrini); and *Xenismarus* (Diapriinae: Spilomicrini).

Methodologically we follow the groundwork provided by Lutzoni et al. (2000) with respect to analyzing variable length regions and maintaining positional homology (partitioning), and Smith et al. (2004) for base-pairing regions. With this basis we describe several new or revised parsimony-based analyses that variously take into account structural elements. To facilitate these methods we more formally describe a simple alignment format that has been used in several of our recent papers (Gillespie et al. 2005abc; Deans et al. 2006; Wharton et al. 2006), and provide a set of script-based utilities that handle this format and the structural properties embedded in it.

Materials and Methods

Alignment

A full review of the methods that produce “structural” alignments is beyond the scope of the present paper (but see Morrison, 2006), and indeed what constitutes such an alignment, or how one is created, is not well defined. Rather than debate the precise definition we highlight four characteristics of alignments (partitions, "alignability", helices, and mask) that incorporate information from structure. These characteristics are typically not found in published multiple-sequence alignments. The format described here is largely a formalization of Kjer's (1995) method for annotation and markup of rRNA alignments. While these four elements have been variously implemented in our and others efforts (Table 5.1), it is our hope that their definition here should allow for more precise methodologies or definitions to evolve. One of the criticisms of Ogden et al. (2005) is that no tests or specific methods were provided by Kjer (2004) for which his claims can be tested. Accordingly, it seems true that those advocating structural alignments for the purposes of phylogenetic analysis require a more formal and more accessible methodology. The elements described here in conjunctions with the Psy software introduced below will provide an impetus for more rigorous tests of hypotheses derived using structural approaches.

Partitions. Our alignments divide sequence data according to a hypothesis of secondary structure, and for a given alignment the same hypothesis is applied to every

sequence. Each partition represents a transition from one identifiable structural element to another, for example from a 5'-strand of a helix to a hairpin-stem loop to a 3'-strand of the same helix. Since there are many structural elements in a typical rRNA molecule, the end alignment has many partitions, in many cases over an order of magnitude more than typically used in analyses to-date (Table 5.1). Whenever possible each partition should be given a label that ties that region to a known structure. Partition boundaries are delimited by whitespaces (Fig. 5.2) in the alignments described here.

"Alignability". The decisions as to whether or not to align a particular region under algorithmic or other criteria can be based on a wide range of criteria (Lee, 2001; and see discussion in Smythe et. al., 2006). All sequence data, regardless of how disparate they are, are algorithmically alignable. That is, while the end product may be meaningless given "unalignable" data (for example two non-homologous genes), it is still possible to apply an algorithm to that data. The term "ambiguously aligned" (Lee, 2001) is commonly used to describe regions of an alignment wherein an observer believes an algorithm to have failed.

When positional homology hypotheses are formed they are based largely on hypothesis of covariation among paring nucleotides (Kjer, 1995). A structural alignment typically excludes a large proportion of the data because there is little biological justification (i.e. hypothesis of covariation cannot be justified) for aligning individual nucleotides in highly variable regions; these typically corresponding to loops or highly variable pairing regions in rRNA molecules. In this light we propose that a better criterion for application of the term "unalignable" is whether or not a structural (i.e. ultimately "biological") model can be used to guide alignment. For the purposes of the data presented here the property of "unalignability" is indicated by the presence of '[]' surrounding a given partition (Figure 5.1). For those not using structural criterion it is possible to exclude data (i.e. hypothesize "unalignability") using algorithmic means, for example through programs such as gBlocks (Castresana, 2000) or SOAP (Löytynoja and Milinkovitch, 2001). These algorithmic approaches introduce another set of parameters that must be optimized, perhaps using sensitivity analyses.

Helices. Both the 3' and 5' regions of a given helix are considered separate partitions given the schema described above. These partitions may be separated from

each other by other structural elements, thus additional information is required to indicate which partitions in the alignment pair. In the model we present here all partitions are indexed from 0 and given a name (Fig. 5.2). If the name differs by a single quote (') (e,g, 2d and 2d' in Fig. 5.2) then those partitions are identified as pairing.

Mask. In the case of rRNA data a majority of positions for the gene in question are involved in baseparing (Table 5.1). Basepairing hypotheses are established with a pairing mask (Fig. 5.2). The mask is variously used in analysis programs such as PHASE (Jow et al., 2002; Hudelot et al., 2003), in the Stockholm format (see http://www.cgb.ki.se/cgb/groups/sonnhammer/Stockholm.html) and modelers such as Infernal (S. Eddy, available at http://infernal.janelia.org/). We use '( )' in conjunction with the helix indexing to identify pairing positions (Fig. 5.2). Using partition indices and the mask in combination allows for the inclusion of pseudoknots (non-nested pairings) in analyses by scripted reordering of the alignment into the nested format required by most analysis tools (e.g. PHASE).

Construction. While there are a number of alignment editors that capture some of the components described above we are not aware of any that *natively* capture the information required to define all the elements listed above. We expect this problem to be resolved shortly, and indeed editor/modelers such as 4SALE (Seibel et al., 2006) are coming very close to being able integrate and report all the elements discussed above. Practical solutions are possible using existing editors, however, for example two-stage workarounds (e.g. replacing symbols in a given alignment with whitespace to introduce the partition boundaries) are possible. Editors that manage some of the elements listed above include RALEE (Griffiths-Jones, 2005), MacGDE (Eric Litton, available at http://www.msu.edu/~lintone/macgde/), and BioEDIT (Tom Hall, available at http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Several older editors such as DCSE (De Rijk and De Wachter, 1993) and AE2 (Macke in Larsen et al., 1993) are apparently no longer actively being developed. Our alignments are edited and compiled in simple text editors. An overview of the base-format is given in Figure 5.2.

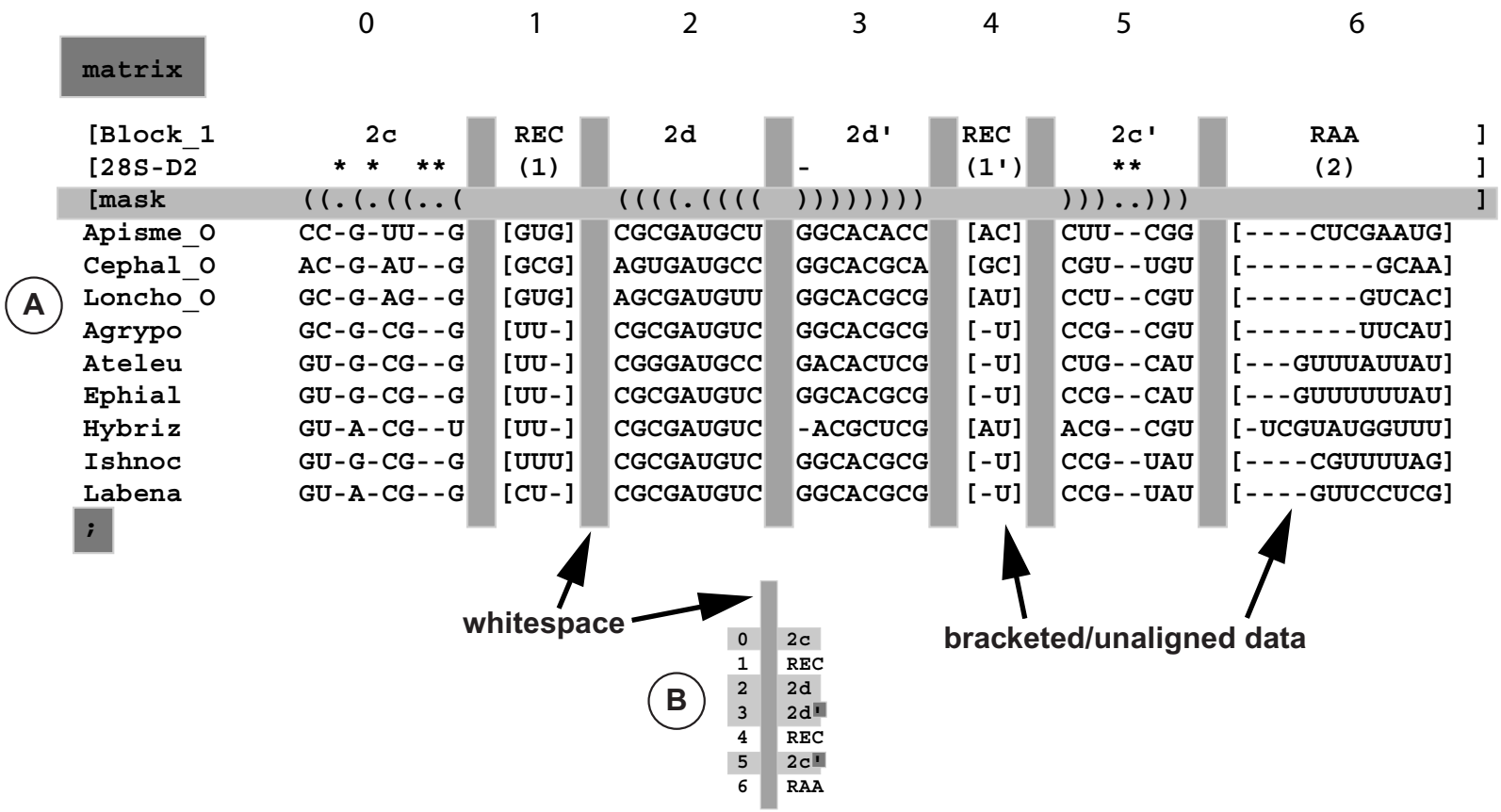

FIGURE 5.2**.** Example format for Psy/jRNA legally formatted files illustrating demarcation of structural elements. The two required text files are A and B. Partitions in file A are delimited by whitespace (spaces, tabs are not allowed). Partitions involved in pairing are defined in file B. Each line in file B is numbered and references a partition. The second column in file B names the partition. If two partitions differ only by a single quote they are considered to be paired (i.e. define a helix). In this example partitions (1,5) and (2,3) form helices. Partitions that are bracketed are considered to be unaligned.

Psy

The "Psy" ('Perl for SYstematics') scripts were created by us to simplify some of the tasks pertaining to the phylogenetic analysis of highly partitioned rRNA datasets. In particular the scripts handle the manipulation and translation of the four elements identified above: 1) the partitions themselves, 2) the property of "alignability", 3) hypotheses of helical regions (i.e. the folding of one partition onto another), and 4) hypotheses of individual positional covariation. These four elements are all incorporated in a modified Nexus legal file format and an additional two-column text file (the 'stem-index') that identifies which partitions pairs to form helices. Psy is an evolution of the jRNA code (MJY, available at http://hymenoptera.tamu.edu/rna), written as a set of Perl modules. The scripts make use of the 'Template' module (available at the Comprehensive Perl Archive Network, CPAN, http://search.cpan.org/dist/Template-Toolkit/lib/Template.pm), which makes the rapid re-composition of input data to numerous output formats possible. For example, POY (Wheeler et al. 2003), PHASE (Jow et al. 2002; Hudelot et al. 2003), Nexus, and Phylip based formats have been created. Where applicable, for a given format, hypotheses pertaining to structure (mask, helical regions) are incorporated into the output. It should be relatively straightforward to write a translation to the RNAML format (Waugh et al., 2002) using the existing code base. Psy is not an alignment algorithm, nor does it contain one. Psy can produce formatted output that is used in Infernal (http://infernal.janelia.org/) or other RNA software requiring the Stockholm format. The scripts also integrate with command-line versions of CLUSTAL and MUSCLE (Edgar, 2004), such that individual partitions can be independently aligned and reintegrated into the matrix in a trivial manner. Psy commands are executed through script files. They are available on request from MJY.

(Hyper) Partitioning

There are various practical and philosophical reasons to employ highly partitioned rRNA (or other) alignments in phylogenetic studies. Partitions as described herein represent extremely specific hypotheses of homology at the structural level (Gillespie 2004; Gillespie et al., 2004). In theory, the specificity of these hypotheses makes them more falsifiable, which is a desirable property. This specificity of these hypotheses also leads to a more meaningful *a-posteriori* examination of the tree-data relationship. For instance, interesting characteristics of structural evolution are able to be more precisely mapped to specific regions of the molecule in question. For rRNA molecules, two different types of homology statements can be derived: those from columns (individual nucleotides), and those delimited by partition. In the latter case the partition is the character and the character states are scored based on properties of the bounded nucleotides. Treatment of partitions as homology hypotheses in this manner is justifiable if indeed evolution acts to conserve structure, not base-composition.

Highly partitioned alignments in which the partitions themselves represent homology hypotheses are practically useful starting points for phylogenetic analyses. In rRNA molecules highly conserved regions are more easily identified, and therefore partitioned, than more variable regions. The boundaries of these regions would typically be embedded in algorithmic multiple sequence alignments. Alignments wherein more explicit hypotheses of homology are made contain these boundaries and typically many additional ones. This is most important because it means that hyper-partitioned alignments can "collapse" to more conservative estimations (e.g. with fewer restrictions imposed by partition boundaries) of primary homology. Collapsed alignments are structural alignments wherein contiguous partitions of similar state (i.e. "alignable" or "unalignable") are fused. For instance, a structural alignment may identify three contiguous helices and separate them into three partitions- these helices all are "alignable" (according to covariation analysis), and so are fused when collapsed. Other algorithmic means to fuse partitions are easily envisioned, perhaps to return partitions of a mean length, or composition. When this functionality is combined with the premise that unalignable regions *should not be aligned by eye* then the arguments of Ogden et al. (2006), who state that manual structural alignments are unrepeatable, become largely

moot. That is, a highly annotated structural alignment can be used to repeat the less-partitioned analyses that practitioners who are concerned about the possible subjective nature of structural alignments are comfortable with. Note that this is only possible if the partitioned alignment is properly (typically highly) annotated.

Partition-based approaches to multiple sequence alignment can result in both more accurate alignments (Morgenstern et al., 2006) and more efficient analyses (e.g. using POY with many small partitions rather than few large ones, see POY manual). Furthermore, hyper-partitioned alignments provide a basis for extremely explicit likelihood-based modeling and metrics like the ILD (see discussion in Dowton and Austin, 2002) or partitioned Bremer support ("PBS") (Baker and DeSalle, 1997; Lambkin et al. 2002).

There are several potential pitfalls with partitioning data. The most obvious arises when a partition inadvertently splits an insertion that was the result of a single mutation event. Subsequent alignment or treatment of the two partitions in this case should have a higher chance of introducing homoplasy as the true bounds of the insertion have been lost. This problem (chance of splitting single mutations) is potentially greatly amplified in hyper-partitioned datasets. By generating data from partitioned and non-partitioned alignments we provide a preliminary test of this problem using the *k*-word method introduced below,

Analyses

A number of likelihood-based approaches that take into account structural information are presently available studies (e.g. Telford et al. 2005). These studies have concluded that accounting for structural information improves estimation of phylogeny (Telford et al. 2005). We chose not to explore these analyses simply because we wished to more fully explore the utility of the parsimony-based approaches developed here. Various other analyses with Psy and our alignments are possible, however, as we have illustrated elsewhere (Gillespie et al. 2005abc; Deans et al. 2006; Wharton et al. 2006). Direct optimization in POY (Wheeler et al., 2003) and the Bayesian criterion in PHASE (Jow et al. 2002; Hudelot et al. 2003) are particularly pertinent to our alignment format.

The parsimony-based analyses used here involve characters that are based on the translation of a single partition or pair of columns into one or more characters. These characters can be used alone or in combination with untranslated characters (standard columnar alignments). Table 5.1 lists the possible combinations we tested, though any possible combination of translated blocks could be attempted. There are three categories of translation types: *k*-word, ARC-based and "paired". The former two act on a given partition, the latter on a pair of columns. A fourth type of analysis simply uses a standard multiple sequence aligner such as CLUSTAL to align the unaligned partitions independently, then returns them to the master alignment. The four analysis categories above can be variously combined into "mixed" analyses (see Table 5.7).

k-*word parsimony*

The ARC package of Kauff et al. (2003) provides a mechanism whereby a user can *a-priori* define a motif whose presence or absence in a given ambiguously aligned partition is then coded for as a single character. The presence or absence of the motif can be found anywhere within the partition, thus the coding mechanism is alignment independent. We extend this idea to find all possible motifs (or length *k* words = "*k*-word", also used variously "*k*-mer", "*k*-String", *k*-tuples) in the partition, and then treat each motif as a separate character whose presence or absence is coded for all taxa in the partition (Figure 5.3). A motif (word) library is found by defining a window of size *k*, then sliding this window along the sequence, moving it forward one nucleotide at a time. Our scripts allow any size or combination of sizes of *k*-word to be defined, and by default only those *k*-words that are parsimony informative are returned, though an option exists to have all for all *k*-words to be returned. Because we find all *k*-words, we eliminate the a-priori nature of deciding which *k*-words to score as characters. Given partitioned data such as ours the method is applicable in two ways. The library of *k*-words to be used can be built and characters scored separately for each individual partition, or partitions can be variously fused (potentially eliminating the complete alignment) and treated as one or more larger partitions. The data are then analyzed under standard unweighted parsimony.

A

```
b    ACGTATGAC
d    ACGTA-CAG
e    ACGTA-CAG
f    A-GTT-GAG
```

B

```
b    ACGTATGAC
d    ACGTACAG
e    ACGTACAG
f    AGTTGAG
```

C

```
b     2   AC CG GT TA AT TG GA AC
      3   ACG CGT GTA TAT ATG TGA GAC
      4   ACGT CGTA GTAT TATG ATGA TGAC
      5   ACGTA CGTAT GTATG TATGA ATGAC
      6   ACGTAT CGTATG GTATGA TATGAC
      7   ACGTATG CGTATGA GTATGAC
      8   ACGTATGA ACGTATGAC
      9   ACGTATGAC
```

D

```
      CA GA TG ACA CAG TAC TGA ACAG GTAC TACA CGTAC GTACA TACAG ACGTAC CGTACA GTACAG ACGTACA CGTACAG ACGTACAG
```

E

```
b     0  1  1  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0
d     1  0  0  1  1  1  0  1  1  1  1  1  1  1  1  1  1  1  1  1
e     1  0  0  1  1  1  0  1  1  1  1  1  1  1  1  1  1  1  1  1
f     0  1  1  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0
```

FIGURE 5.3. An illustrations of the *k*-word method used here. Partitions (e.g. A), if aligned are stripped of their alignment (resulting in B), or this stage is skipped for unaligned data. A library of all strings of a given length (in this example all possible lengths) is built . An example of all possible words of all lengths for sequence 'b' is provided in C. Words from all sequences are concatanted to the same library (*i.e.* repeat C for sequences d, e, f). Only those words that are found in two or more taxa are kept (in this example the result is D). The presence or absence of these words (e.g. "TGA"), regardless of where they are found in the sequence, is coded for each sequence (as seen in E). This matrix is then analyzed under the parsimony criterion.

It is possible to find and then analyze as a combined dataset all possible words of all lengths for a given set of partitions. This is a type of "elision" approach sensu Wheeler et al. (1995). This method suggests that all alignments generated from an "envelope" of parameter costs be generated and analyzed simultaneously. The hypothesis is that signal from sets of analyses of different parameter values will overlap and amplify, while noise will be down-weighted since different parameter values will return varying results. In the case of *k*-word parsimony it is not alignments that are generated, but datasets of different word size, however, the idea remains the same. If certain evolutionary events introduce words of a given size then finding all possible words should have a greater possibility of properly "scoring" these events.

As reviewed by Vinga and Almeida (2003) *k*-word approaches are used in various alignment free approaches to sequence comparison. In these approaches it is primarily the frequency of words that is employed as the base data (Vinga and Almeida, 2003). These comparisons are not always specifically aimed at deriving phylogenies, for example they may be used to hypothesize functional or structural analogs, be applied to chaos game representation (Deschavanne, 1999) or be used as the basis for guide trees for multiple sequence alignment as in MUSCLE (Edgar, 2004).

Various methods have recently employed *k*-words in phylogenetic inference, however, following the construction of the *k*-word library, these methods use distance or likelihood-based approaches on the data (e.g. Pride et al., 2003; Qi et al., 2004ab; Stuart and Berry, 2004; Chapus et al., 2005; Ulitsky et al., 2006; and see reviews in Philippe et al., 2005). These methods have also tended to be focused at genome (e.g. Qi et al., 2004b; Wu and Lin, 2005) or very large-scale datasets (e.g. whole mitochondria), rather than at small partitions or numbers of genes (though see Chapus et al. 2005). Many of these methods also use a fixed window size (*k*-word length) for statistical or model based-reasons (Vinga and Almeida, 2003). The parsimony-based approach as presented here may be less applicable at genome-sized scales for large ranges of *k*-word sizes because of the large (perhaps intractable) number of characters that would be generated. Employing a suffix-tree based approach (e.g. Angelov et al., 2006) would increase the efficiency of character translation for larger sized datasets.

A related approach has been suggested by Zaldivar-Riverón et al. (2006) wherein fragments of the same length from ambiguously aligned partitions are extracted and analyzed as if aligned. This approach employs a window, but treats the data of the same window length as aligned.

We have not fully explored the justification to the parsimony approach detailed here, rather we sought to first empirically test its potential utility. One immediately recognizable critique is that characters generated by the method are potentially non-independent. For example, shorter length words (e.g. 2-4) will be relatively frequent, and since only their presence or absence anywhere in the partitioned string is coded, the probability of violating positional homology increases. The non-independence should be reduced given hyper-partitioned alignment, as characters are only coded within partitions. To what extent this is the case, however, remains to be seen. Theoretical background for distance and likelihood based methods that employ *k*-words is building (e.g. Vinga and Almeida, 2003; Qi et al., 2004b; Chapus et al., 2005; Ulitsky et al., 2006; and in particular see Wu et al., 2005), and may ultimately be applicable to the understanding of how these parsimony based methods work, but most acknowledge that more work is needed to understand the full implications of these methods.

Various statistics or weighting schemes can be envisioned as additions to the basic *k*-word method. An ILD ("kILD") can be calculated for analyses involving a single partition, and multiple *k*-word sizes following the same formula as that for the ILD sensu Wheeler (1995) (see Dowton and Austin, 2002 for discussion of other ILD metrics):

$$\frac{(\text{Sum scores each k - word combined}) - (\text{Sum scores each k - word alone})}{\text{Sum scores each k - word combined}}$$

Congruence increases as kILD decreases. One possible means to employ this test is to run a set of analyses alone and in combination, for example for word sizes 3-8 (an arbitrarily selected range). The kILD can then be calculated for all sets of word sizes (e.g. [3,4], [3,4,5], [3,4,5,6] … [4,5], [4,5,6] … [7,8]) to determine the effect of combining *k*-word derived datasets.

We have not explored here weighting schemes for *k*-word based characters. This approach is clearly possible, down-weighting more commonly encountered or more statistically common *k*-words. For the combined *k*-word approach (i.e. analyzing simultaneously recoded data for two or more word lengths) a simple weighting scheme could be down-weighting a given word size proportional to the longest length. For example given a set of lengths [3,4,5,6,7,8] length 3 words would be weighted 1/6, length 7 5/6, etc. This scheme would have the effect of decreasing the number of state-transitions for long words (those that are probably less homoplasious). Other weighting schemes could be derived using an information-theoretic background; in particular the work of Wu et al. (2005) should provide a basis from which to derive different approaches, and see Kjer (1995) for comments on weighting and references therein.

The *k*-word method outlined here presents several attractive qualities. As noted by Qi et al. (2004b) the number of free parameters involved in such analyses is drastically reduced, in our case essentially to zero when all possible lengths of *k*-word are found and included. That is, no gap-opening or extension costs (values with little if any biological meaning, Gusfield, 1997) or model parameters need to be used (Chapus et al., 2005). The method itself is very parsimonious in this sense. In our experience the method (i.e. using a single word length) can be rapidly applied for small (~20 terminals) to medium (~300 terminals) sized datasets of moderate length (~500-800 nucleotides).

Partition based characters

Miadlikowska et al. (2003), implemented in Kauff et al. (2003), proposed a method of recoding variable regions into a set of 24 descriptive characters, which where then further translated into fixed states for parsimony analysis.implementation in PAUP*. This was accomplished by the "ARC" (Ambiguous Regions Coding) scripts. One class of these characters (see also description of the *k*-word parsimony method) was essentially a set of ratios (Table 5.2), for example the percentage of given nucleotide for a sequence. Each unique ratio was given a separate state, and if too many states occurred (effectively 63 in PAUP*) the ambiguous region could not be coded for that ratio. This method has been translated into the Psy package, with several modifications. In addition to including

the original ARC functionality, Psy provides functionality for leaving the states as real numbers (primarily ranging from 0 to 1.0). These characters are then analyzable in TNT as continuous (ordered) characters. This method allows for the use of many more states (64,000), at the limitation that they must be treated as ordered characters. Several descriptive characters have been added to the method of Kauff et al. (2003), and the system may be further expanded in a relatively easy manner to include any conceivable character recoding. Application of the method can be made to multiple partitions (or fusion thereof) by Psy's partition-handling capabilities (it is also possible to code multiple regions independently with ARC). To differentiate our modification we use the name ARCO (for "ordered").

TABLE 5.2. Characters modified from the ARC suite of Kauff et al. (2003). The variables n_x and n_y can be any nucleotide A, C, G or T; n_{acgt} represents the total of all nucleotides. Characters 23 and 24 are not employed in Kauff et al. (2003) or in the analyses herein (they are available in Psy). They are included to illustrate additional potential codings.

#	Short name	Calculation
1-4	pct_[acgt]	Total n_x / Total n_{acgt}
5-14	pct_[aa, ac, ag, at, cc, cg, ct, gg, gt, tt]	Total n_xn_y / Total n_{acgt} - 1
15-18	pct_[acgt]_pair	Total n_xn_x / Total n_{acgt}
19-22	pct_[acgt]_dist	Total nucleotides between n_x / Total n_x (e.g., for case c: `cttc` = 1/2; `tcttaca` = 2/3; `cttcttca` = 3/4; `atgcatg` = 0; `catg` = 0)
23	blk_length	Total n_{acgt}
24	presence	1 - Any n_x present; 0 - no nucleotides present

Paired-column characters

It has long been recognized that pairing regions in rRNA molecules evolve non-independently (Wheeler and Honeycutt , 1988; Lindgreen et al. 2006). Honeycutt and Wheeler (1988) note: "When characters are functionally linked or part of some ontogenetic complex, they are only expressing a single source of information. Only independently varying qualities are viable characters in phylogenetic reconstruction." Covariation hypotheses are most simply implemented into phylogenetic analyses using a mask statement (e.g. the "#=GC SS_cons" line of Stockholm formatted files). While this implementation (i.e. the process of creating an input file with a mask) is technically

simple, the process of deciding which sites covary is non-trivial (see Lindgreen et al. 2006 for a recent overview). We do not wish to argue the relative merits of the various methods that generate mask statements, but rather to explore what can be done with an alignment annotated by a pairing mask.

Two, perhaps extreme, means of removing non-independence from analyses that require independence are to merge two non-independent sites into one or to eliminate one of the two sites. Translation of two sites into one has been implemented in a likelihood framework by Smith et al. (2004). This same approach is also possible in a parsimony-based framework, though we are unaware of any existing implementations. Our scripts include a conversion module that translates the columns involved in pairing into single columns following the convention of Smith et al. (2004) (see Table 5.3). The scripts also allow either the 3' or 5' sides of the stems to be isolated and analyzed alone.

TABLE 5.3. Translation table for nucleotides involved in basepairing. Modified from Smith et al. (2004). All nucleotides involved in helices (left columns) are translated to a new alphabet (right column). 'X' indicates any other non-ACGU character, e.g. IUPAC codes.

A A	CA H	A- A	XA -
C R	CC I	-A A	AX -
G N	CG L	C- R	UX -
U D	CU K	-C R	XU -
AA C	GA M	G- N	XG -
AC Q	GC F	-G N	GX -
AG E	GG P	U- D	XC -
AU G	GU S	-U D	CX -

The translated columns can then be analyzed "as-is" or with weighting schemes, perhaps based on a covariation measure (e.g. Lindgreen et al., 2006) For example, characters could be weighted proportional to their Mutal Information (MI, see Lindgreen et al. 2006), or filters that exclude columns not significantly covarying can be generated. We implemented weighting schemes based on the standard MI , Cramer's V statistic (Cramér, 1999), and an inclusion/exclusion filter based on a Chi^2 test of covariation. Our initial exploration of these approaches (data not shown) provided ambiguous results, i.e. it is unclear how they affected phylogeny or whether they are warranted. It is possible, however, that future refinement of weighting schemes, perhaps using the various MI

derivations of Lindgreen et al. (2006) may be of use. We believe further exploration is warranted.

Diapriid data

We use the methods described here to analyze a dataset consisting of 28S and 18S rRNA data for 168 taxa of the family Diapriidae (Insecta: Hymenoptera). The dataset is in active development, and as such contains missing data (Tables 5.4, 5.5). Nevertheless, we have successfully sampled a broad spectrum of diversity for the family for a large number of taxa. In addition, for 10 of the taxa (Table 5.5) we sequenced the majority of the 28S rRNA and all the 18S rRNA. Four additional outgroups (*Ateleute*, *Labena*, *Megalohelcon* and *Thoracoplities*) from the Ichneumonoidea were taken from Belshaw and Quicke (2002). The core matrix presented here contains 161 28S rRNA D2, 136 28S rRNA D3-D5, and 97 18S rRNA V4 amplicons. Both core and expanded data are all included in the global alignment. To minimize the effects of missing data we variously eliminated taxa, primarily under the criterion that taxa contain > 20% of the characters for a given analysis. We selected taxa specifically to address two major questions: 1) are the four diapriid subfamilies as presently defined monophyletic and 2) what are the relationships among them?

TABLE 5.4. Included taxa and rRNA gene regions. A sequence ID in the "D2" (= 28S D2), "D3-5" (=28S D3-D5) or "18S" (=18S V4) columns indicates the presence of that amplicon. Sequences were stored in an instance of the mx database (see links from http://www.diapriid.org) and were attached to the OTU ID listed in column "OTU". Where Taxon is not a genus group name the taxon is undescribed. Where indicated by [1] or [2] sequences were provided by the John Heraty Lab, (UCR) and Debra Murray, (FSU, HymATOL) respectively. Several sequences were included from Genbank. This list does not include the taxa used in Deans et al. (2006). See also Table 5.5.

Family	Subfamily	Taxon	OTU	D2	D3-5	18S
Ceraphronidae		*Aphanogmus* [1]	1689	722		
Ceraphronidae		*Ceraphron* [1]	1690	723		769
Diapriidae	Ambositrinae	*Ambositra famosa*	1212	944	946	
Diapriidae	Ambositrinae	*Ambositra famosa*	1206	573	574	
Diapriidae	Ambositrinae	*Ambositra famosa*	1293	983	559	653
Diapriidae	Ambositrinae	Ambositrinae	422	890	652	689
Diapriidae	Ambositrinae	Ambositrinae	404	608		677
Diapriidae	Ambositrinae	Ambositrinae	413	879		682
Diapriidae	Ambositrinae	Ambositrinae	434	901	628	
Diapriidae	Ambositrinae	Ambositrinae	1268	978	553	642
Diapriidae	Ambositrinae	Ambositrinae	399		859	
Diapriidae	Ambositrinae	Ambositrinae	391	849		
Diapriidae	Ambositrinae	*Austroxylabis pictipennis*	408	869		679
Diapriidae	Ambositrinae	*Diphoropria*	1233	969	970	971
Diapriidae	Ambositrinae	*Diphoropria*	411	875	876	
Diapriidae	Ambositrinae	*Diphoropria*	412	877	878	681
Diapriidae	Ambositrinae	*Diphoropria*	1661	701		
Diapriidae	Ambositrinae	*Diphoropria*	1665	705		
Diapriidae	Ambositrinae	*Fanis*	400	606	860	673
Diapriidae	Ambositrinae	*Gwaihiria*	410	873		
Diapriidae	Ambositrinae	*Gwaihiria*	1277	1123		
Diapriidae	Ambositrinae	*Pantolytomyia*	1274	1113	1114	1112
Diapriidae	Belytinae	*Aclista*	1147	920	544	666
Diapriidae	Belytinae	*Aclista*	1223	589	590	694
Diapriidae	Belytinae	*Aclista*	429		624	
Diapriidae	Belytinae	*Aclista*	1667	707		
Diapriidae	Belytinae	*Aclista*	1668	708		
Diapriidae	Belytinae	*Aclista* [1]	1679	713		759
Diapriidae	Belytinae	*Acropiesta*	1135	909	532	
Diapriidae	Belytinae	*Acropiesta*	1145	917	918	665
Diapriidae	Belytinae	*Anommatium* [1]	1683	717		763
Diapriidae	Belytinae	*Belyta*	1150		922	923
Diapriidae	Belytinae	*Belyta*	1204	939	571	
Diapriidae	Belytinae	*Belyta*	1278	1125	1126	
Diapriidae	Belytinae	*Belyta*	1151	925		668
Diapriidae	Belytinae	Belytinae	415	882	647	683
Diapriidae	Belytinae	Belytinae	416	883	648	684
Diapriidae	Belytinae	Belytinae	388	557	848	

Table 5.4 Continued.

Family	Subfamily	Taxon	OTU	D2	D3-5	18S
Diapriidae	Belytinae	Belytinae	420	887	651	
Diapriidae	Belytinae	Belytinae	432	899		
Diapriidae	Belytinae	Belytinae	1660	700		
Diapriidae	Belytinae	*Camptopsilus*	401	863		674
Diapriidae	Belytinae	*Cinetus*	1200	936	935	
Diapriidae	Belytinae	*Cinetus*	1222	959	588	657
Diapriidae	Belytinae	*Cinetus* [1]	1681	715		761
Diapriidae	Belytinae	*Eccinetus*	1195	930	929	
Diapriidae	Belytinae	*Eccinetus*	1208	576	577	
Diapriidae	Belytinae	*Gladicauda*	406	871	870	
Diapriidae	Belytinae	*Lyteba*	425	891	622	
Diapriidae	Belytinae	*Lyteba*	428	894		
Diapriidae	Belytinae	*Masnerosema*	1137	912	534	911
Diapriidae	Belytinae	*Masnerosema*	1213	581	582	
Diapriidae	Belytinae	*Miota*	1221	587	958	957
Diapriidae	Belytinae	*Miota* [1]	1680	714		760
Diapriidae	Belytinae	*Opazon* [1]	1682	716		762
Diapriidae	Belytinae	*Pantoclis*	430	896	625	
Diapriidae	Belytinae	*Pantoclis*	1152	927	547	669
Diapriidae	Belytinae	*Pantoclis*	1207	575	942	
Diapriidae	Belytinae	*Pantoclis*	1224	960	591	658
Diapriidae	Belytinae	*Pantoclis*	1225	961	962	659
Diapriidae	Belytinae	*Polypeza*	1136	910	533	637
Diapriidae	Belytinae	*Stylaclista*	1232	527	528	968
Diapriidae	Belytinae	*Stylaclista*	1227	964	592	660
Diapriidae	Belytinae	*Synbelyta*	1201	937	569	
Diapriidae	Diapriinae	*Acanthopria*	1205	940	572	654
Diapriidae	Diapriinae	*Alareka*	1210	579	945	655
Diapriidae	Diapriinae	*Aneurhynchus*	1144	542	541	
Diapriidae	Diapriinae	*Aneurhynchus*	1276		1120	
Diapriidae	Diapriinae	*Aneurhynchus*	1219	586	954	692
Diapriidae	Diapriinae	*Aneurhynchus* [1]	1684	718		764
Diapriidae	Diapriinae	*Basalys*	1142		914	638
Diapriidae	Diapriinae	*Basalys*	1264	973	550	670
Diapriidae	Diapriinae	*Basalys*	1202	938	570	
Diapriidae	Diapriinae	*Basalys*	1194	564	928	
Diapriidae	Diapriinae	*Basalys*	1220	955	956	693
Diapriidae	Diapriinae	*Basalys*	1666	706		
Diapriidae	Diapriinae	*Calogalesus*	1217	950	951	
Diapriidae	Diapriinae	*Calogalesus*	1216	585	949	
Diapriidae	Diapriinae	*Chilomicrus*	419	620	650	687
Diapriidae	Diapriinae	*Chilomicrus*	402	865	866	675
Diapriidae	Diapriinae	*Coecopria*	1265	974	975	
Diapriidae	Diapriinae	*Coptera*	1218	953	952	
Diapriidae	Diapriinae	*Coptera*	1296	987	988	986
Diapriidae	Diapriinae	*Coptera*	1297	989	562	
Diapriidae	Diapriinae	*Coptera*	1196	931	565	
Diapriidae	Diapriinae	*Diapriinae*	1229	523	966	634

Table 5.4 Continued.

Family	Subfamily	Taxon	OTU	D2	D3-5	18S
Diapriidae	Diapriinae	*Diapriinae*	433	900	627	
Diapriidae	Diapriinae	*Diapriinae*	392	851		
Diapriidae	Diapriinae	*Diapriinae*	1139	537	536	
Diapriidae	Diapriinae	*Diapriinae*	1140	539	538	
Diapriidae	Diapriinae	*Diapriinae*	1138	535	913	
Diapriidae	Diapriinae	*Diapriinae*	403	607		676
Diapriidae	Diapriinae	*Diapriinae*	414	881	646	
Diapriidae	Diapriinae	*Doddius*	1266	551	976	640
Diapriidae	Diapriinae	*Doliopria*	1131	903	904	635
Diapriidae	Diapriinae	*Entomacis*	431	897	626	
Diapriidae	Diapriinae	*Entomacis*	1197	932	566	
Diapriidae	Diapriinae	*Entomacis*	1198	568	567	
Diapriidae	Diapriinae	*Entomacis*	1209	941	578	690
Diapriidae	Diapriinae	*Entomacis*	1275	1117		1116
Diapriidae	Diapriinae	*Idiotypa*	393	861		854
Diapriidae	Diapriinae	*Idiotypa*	1267	977	552	641
Diapriidae	Diapriinae	*Labidopria*	1271	555	556	981
Diapriidae	Diapriinae	*Leucopria*	1235	530		663
Diapriidae	Diapriinae	*Megaplastopria*	1134	907	531	636
Diapriidae	Diapriinae	*Neurogalesus*	418	886	649	686
Diapriidae	Diapriinae	*Neurogalesus*	1228	522	965	633
Diapriidae	Diapriinae	*Neurogalesus*	1230	524	525	661
Diapriidae	Diapriinae	*Neurogalesus*	1231	967	526	662
Diapriidae	Diapriinae	*Paramesius*	1133	906	905	664
Diapriidae	Diapriinae	*Paramesius*	1215	948	584	
Diapriidae	Diapriinae	*Pentapria*	387	846	847	643
Diapriidae	Diapriinae	*Pentapria*	1269	980	554	979
Diapriidae	Diapriinae	*Pentapria*	417	885	884	685
Diapriidae	Diapriinae	*Pentapria*	397	605	645	
Diapriidae	Diapriinae	*Poecilopsilus*	386	844	845	671
Diapriidae	Diapriinae	*Psilus*	1292	982		
Diapriidae	Diapriinae	*Spilomicrus*	1663	703		
Diapriidae	Diapriinae	*Spilomicrus*	1662	702		
Diapriidae	Diapriinae	*Spilomicrus*	1664	704		
Diapriidae	Diapriinae	*Spilomicrus*	1199	934	933	
Diapriidae	Diapriinae	*Spilomicrus*	421	889	888	688
Diapriidae	Diapriinae	*Trichopria*	1294	984	560	
Diapriidae	Diapriinae	*Trichopria*	1295	985	561	
Diapriidae	Diapriinae	*Trichopria*	1148	921	545	667
Diapriidae	Diapriinae	*Trichopria*	1298	990	563	
Diapriidae	Diapriinae	*Trichopria*	1214	947	583	691
Diapriidae	Diapriinae	*Trichopria*	1211	943	580	656
Diapriidae	Diapriinae	*Xenismarus*	398	857		672
Diapriidae	Ismarinae	*Ismarus* [2]	1248	696	697	695
Figitidae		*Paraspicera* [1]	1686	720		766
Heloridae		*Helorus*	395	855		
Ibaliidae		*Ibalia*	1143	916	540	
Liopteridae	Mayrellinae	*Paramblynotus* [1]	1685	719		765

Table 5.4 Continued.

Family	Subfamily	Taxon	OTU	D2	D3-5	18S
Maamingidae		*Maaminga rangi*	1659	709		
Megaspilidae	Megaspilinae	*Dendrocerus* [1]	1688	721		767
Monomachidae		*Monomachus*	1171	1109	1110	1108
Monomachidae		*Monomachus* [1]	1670	711		
Monomachidae		*Monomachus antipodalis*	1669	710		
Mymarommatidae		*Palaeomymar* [1]	1698	731		777
Mymarommatidae		*Palaeomymar* [1]	1701	734		780
Mymarommatidae		*Palaeomymar* [1]	1700	733		779
Mymarommatidae		*Palaeomymar* [1]	1699	732		
Platygastridae	Platygastrinae	*Synopeas* [1]	1697	730		776
Platygastridae	Sceliotrachelinae	*Aphanomerus* [1]	1696	729		775
Proctotrupidae	Proctotrupinae	*Exallonyx* [1]	1691	724		793
Scelionidae	Scelioninae	*Archaeoteleia* [1]	1695	728		774
Scelionidae	Scelioninae	*Archaeoteleia* [1]	1694	727		773
Scelionidae	Scelioninae	*Archaeoteleia* [1]	1693	726		772
Scelionidae	Scelioninae	*Macroteleia* [1]	1692	725		771

TABLE 5.5. Taxa sequenced for the complete 18S and 28S rRNA.

				18S			28S						
Family	Subfamily	Taxon	OTU	1	2	3	D1	D2	D3-5	D6-D7	D8-D9	D10	D12
Diapriidae	Ambositrinae	Ambositrinae	405	609	807	610	820	832	827	611	612	831	1104
Diapriidae	Ambositrinae	*Pantolytomyia*	407	613	678	614		824	825	615	821	822	1105
Diapriidae	Diapriinae	*Aneurhynchus*	1149	595	812	596	835	836	546	837		597	1100
Diapriidae	Diapriinae	*Entomacis*	1226	629	813	630	631	963	839	632	840	841	1107
Diapriidae	Diapriinae	*Idiotypa*	389	804	644	599	816	558	817	600	818	601	1102
Diapriidae	Belytinae	*Belyta*	409	616	680	811	618	617	826	833	619	834	1106
Diapriidae	Belytinae	Belytinae	390	805	806	602	603	829	828	604	819	830	1103
Monomachidae		*Monomachus*	1272	814	815	598	1170	842	843	1169	1167	1168	1101
Proctotrupidae	Proctotrupinae	*Disogmus*	1146	593	639	594	1165	919	543	1164	1163	1166	1099

Diapriids are cosmopolitan, and over 4000 species are estimated to exist. Taxon sampling to adequately represent the diversity in the family is therefore non-trivial. We succeeded in sampling a large number of genera from most of the major lineages hypothesized from morphological data. Of these groups, selection is biased towards Southern hemisphere taxa, as many purportedly primitive lineages have an apparent Gondwanian distribution. As might be expected, a large number South American diapriids remain undescribed, particularly in the most taxonomically poorly understood subfamily, the Belytinae. While it is typically undesirable to include undescribed taxa, in the case of the Belytinae, where generic limits are poorly understood, any insights that will help define relationships and monophyletic clades are desirable. Diapriidae also contains one monomorphic, enigmatic, subfamily, the Ismarinae. Species of *Ismarus* have, to our knowledge, never been included in published phylogenetic analyses. Their morphology is distinct enough to warrant testing of their placement within the Diapriidae.

The placement of diapriids in the large, paraphyletic "proctotrupomorpha" is of great interest to deep-level hymenopteran phylogenetics. A recent grant (NSF HymATOL) is aimed at resolving these deep level questions. In lieu of those results, the outgroups we used are largely members of the Monomachidae and Maamingidae, chosen based on results of Dowton and Austin (2001), Castro and Dowton (2006) and the hypotheses of Early et. al. (2001). It is possible, though perhaps unlikely, that Ismarinae do not form a monophyletic group with Monomachidae+Maamingidae+Diapriidae. To take this into account we included additional outgroups from the Proctotrupomorpha.

Vouchers for the analyses are deposited at the Texas A&M University Insect collection (TAMUIC). Sequence and specimen data were managed in an installation of mx . Supporting data are available at http://www.diapriid.org and following links therein.

Sequencing protocols follow Gillespie et al. (2005c). Primers used are listed in Table 5.6.

TABLE 5.6. Primers used in both PCR and cycle sequencing. See also Gillespie et al. (2005).

Name	Gene	Primer
18S 1F	18S	TAC CTG GTT GAT CCT GCC AGT AG
18S 4R	18S	G AAT TAC CGC GGC TGC TGG
18S-H17 F	18S	AAA TTA CCC ACT CCC GGC A
18S-H35 R	18S	TGG TGA GGT TTC CCG TGT T
18S a2.0 F	18S	ATG GTT GCA AAG CTG AAA C
18S 9R	18S	GAT CCT TCC GCA GGT TCA CCT AC
D1-3317 F	28S	ACC CGC TGA ATT TAA GCA TAT
D2-3551 F	28S	CGT GTT GCT TGA TAG TGC AGC
D2-4068 R	28S	TTG GTC CGT GTT TCA AGA CGG G
D3-4046 F	28S	GAC CCG TCT TGA AAC ACG GA
D5-4625 R	28S	CCC ACA GCG CCA GTT CTG CTT ACC
D4-4410 F	28S	CCG AAG TTT CCC TCA GGA TAG CT
D5-4749 R	28S	GTT ACA CAC TCC TTA GCG GA
D6-4738 F	28S	GGA GTG TGT AAC AAC TCA CCT GCC G
D7-5482 R	28S	CCT TAT CCC GAA GTT ACG
D8-5435 F1	28S	CCC ATA TCC GCA GCA GGT CTC C
D8-5999 R	28S	GGT TTC GCT GGA TAG TAG
D8-5982 F2	28S	CTA CTA (T/A)CT AGC GAA ACC
D10-6582 R	28S	GAA GAG CCG ACA TCG AAG
D11-6389 F	28S	GGA CAT TGC CAG GTA GGG AGT T
D12-7200 R	28S	GCA AAG GAT AAG CTT CAG TGG

Alignment. Data were aligned by JJG who had no taxonomic experience with diapriids. A portion of the original data was labeled only with identifying codes. This process should minimize any potential bias in the alignment by eliminating *a-priori* knowledge of taxa and relationships. The method of alignment is the same as used in Gillespie et. al. (2005abc) and Deans et. al. (2006). The whole alignment consists of 987 partitions, 933 were unbracketed (aligned) and 54 bracketed (unaligned). The total length of characters is 7089, the longest bracketed block 160 characters long. There are 1705 paired columns identified in 249 helicies. The alignment can be browsed at http://hymenoptera.tamu.edu/rna.

Phylogenetic Analyses

The analyses presented here are an empirical test of the methods discussed above using the diapriid alignment. We were broadly interested in examining the following:

- The general functionality of the recoding methods on bracketed (unaligned) data alone. Variable length regions or ambiguously coded data may possess valuable phylogenetic signal (Lutzoni et al., 2000). As hyper-partitioned datasets very explicitly delimit these regions, it should be possible to test this hypothesis.
- The application of recoding methods to both aligned and unaligned data. The homology statements in this case are block-level, i.e. there are no columnar homology statements. If the recoding methodology is reasonable (this may not necessarily be the case), and if the homology hypotheses represented by the partitions in fact reflects evolutionary bounds of structural constraints, then results from this approach should be largely congruent with other sources of data (e.g. morphology), or traditionally robust results (parsimony on aligned data).
- The utility of combining aligned (unbracketed) and recoded (on bracketed partition and/or mask pairs) data.
- The possible consequences of hyper-partitioning multiple sequence alignments.

As a baseline we generated a strict consensus for all taxa and all available data (minus "tails") for aligned (unbracketed) data only (analysis A1, see Tables 5.5, 5.6 to interpret this and following analysis labels). This analysis is extremely conservative, as a large fraction of the alignment is explicitly excluded. The presence or absence of these clades was noted in subsequent analyses. As a further baseline we also generated an algorithmic alignment using L-INS-I strategy in MAFFT (Katoh et al., 2002, 2005). The parameters for these alignments (all defaults or recommended by the Katoh et al. (2002, 2005)) were " --algq --localpair --maxiterate 1000". The "--algq --localpair" options are experimental in the version of MAFFT used, but are noted (Katoh et al. 2002, 2005) as increasing accuracy for sequences with multiple variable regions. MAFFT uses an iterative

algorithm to improve the alignment after its initial build, in this case the maximum number of improvement iterations is set to 1000.

TABLE 5.7. An overview of possible parsimony-based analysis combinations. This list does not including combinations using direct-optimization. Analyses are coded by letter (column 1), and a digit (Table 5.8, "code" column). For example, analysis 6R is the 28S D2 data with both the bracketed and unbracketed data recoded as *k*-words. See Materials and Methods-"Analyses" for an explanation of the "Type" column. "Unbracketed" and "Bracketed" refer to how the aligned and unaligned partition are respectively treated. "Homology type" indicates whether characters are traditional nucleotide/column based (="columnar") or partition-based translations (="block"), or both.

Code	Type	Unbracketed	Bracketed	Homology type
A	standard	standard	excluded	columnar
B	standard	3' only[1]	excluded	columnar
C	standard	5' only[1]	excluded	columnar
E	standard	standard	algorithmically aligned[1]	columnar
F	paired	paired		columnar
J	mixed	paired	ARC[4]	columnar/block
K	mixed	paired	kword	columnar/block
L	mixed	paired	algorithmically aligned[2]	columnar/block
M	mixed	standard	ARC[4]	columnar/block
N	mixed	excluded	ARC[4]	columnar/block
O	ARC[4]	ARC[4]	ARC[4]	block
P	mixed	standard	kword	block/columnar
Q	mixed	excluded	kword	block/columnar
R	kword	kword	kword	block
S	mixed	kword	ARC[4]	block
T	mixed	ARC[4]	kword	block
U	kword	combined[3]		block
V	ARC[4]	combined[3]		block

[1] Includes both pairing and nonpairing positions (e.g. bulges) in partition which bound the helix.
[2] For example by Clustal.
[3] Both unbracketed and bracketed combined and gaps removed prior to further analysis.
[4] Datasets filenames that used the original ARC coding are appended with an additional "o", all others use continuous characters.

TABLE 5.8. Combinations of data partitions explored. All available diapriids are initially included in addition to all outgroups (column 2) or only the monomachoid outgroups (column 3), which includes the members of Monomachidae and Maamingidae. In most cases a filter was additionally run to eliminate taxa with less than 20% sequence coverage. The "code" column is used in the naming convention of analyses, for additional information see explanation in Table 5.7.

Code	Outgroup	"monomachoids"	18S	28S D2	28S D3-D5	Tails[3]		
1	x	x	x	x	x			
2		x	x	x	x			
3	x	x	x					
4		x	x					
5	x	x		x				
6		x		x				
7	x	x			x			
8		x			x			
9	x	x		x	x			
10		x		x	x			
11	x	x	x	x				
12		x	x	x				
13	x	x	x	x	x	x		
14		x	x	x	x	x		
				18S core[2]	18S variable[2]	28S core[2]	28S variable[2]	
15		x	a[1]	x	x			
16		x	a[1]			x	x	
17		x	a[1]	x		x		
18		x	a[1]		x		x	

[1] When present in a filename (i.e. in a code representing an analysis) the additional 'a' indicates that the "full" 18S and 28S data were used, when absent only the D1-D2, D3-D5 and V4 bounding amplicons were used.
[2] For these analyses data were collapsed into two categories, "core" and "variable". These partitions were used in conjunction with U and V codes of Table 5.7.
[3] This partition represents 28S minus D2-D5 and 18S minus the V4 amplicon, it was sequenced for 10 taxa (Table 5.5) as an exploratory exercise.

Early tests (not shown) of the *k*-word method suggested to us that the method was extremely sensitive to missing data. This is evident in that missing data are necessarily coded as 0, and taxa with shared missing data are therefore artificially grouped by false synapomorphies. We therefor focused on the D2 amplicon, for which we had the most data. Analyses followed two courses. In the first we looked at the effect of combining *k*-

word coded characters of various lengths in a single analysis, in the second we analyzed a single length alone. These two approaches were applied in several ways: 1) on the bracketed data alone; 2) on the bracketed alone but analyzed in combination with unbracketed (aligned) data; 3) on all partitions, recoding them all individually; and 4) on all partitions, fusing them (removing the alignment), and treating them as a single partition. For the fixed-length alone analyses we analyzed words of length 3 to 15 (there were no parsimony informative words of length 2). Based on our results (see discussion) for 28S D2 we further analyzed words of length 8 for D3-D5 and 18S data.

To provide an overview of the data generated by the process of *k*-word translation we plotted for each of the three amplicons (28S D2, 28S D3-D5, 18S V4), and for each variable region in those amplicons the number of parsimony informative words for a range of word sizes. We further illustrated several of the matrices generated from the process as bitmap plots.

Our approach to the ARCO recoding method was similar to that of the *k*-word. Our first analyses (not shown) found that treating the characters coded as 0.0 (essentially missing data) as missing ("?") mitigated the effects of missing data. We applied the recoding method on 1) bracketed data alone; 2) bracketed data recoded with unbracketed data included; and 3) all partitions recoded. This was done for partitions coded 1 and 2 in Table 5.8.

The paired-parsimony approach was also used for partitions coded 1,2,4, and 10 in Table 5.8. In these cases we excluded the bracketed data. Characters in these analyses were all treated as unweighted and unordered.

After examining the results of the three methods described above we selected several of those that performed well (as compared with expectations based on the taxonomic classifications of Masner (1993), Naumann (1982), Masner and García (2002) and the analyses performed using the algorithmically aligned data) and combined those approaches. We were particularly interested in whether recoding approaches, which treated all data (both bracketed and unbracketed partitions), could recover results similar to the more traditional approaches (alignment of conserved regions with exclusion of ambiguously aligned data).

Finally, following Müller and Reisz (2006), we used the Bayesian approach of Lewis (2001) on a *k*-word dataset (6U for *k*-word size 8). Analyses were performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). We experimented with rates categories modeled using a gamma distribution (4 categories) and a single substitution rate. Two analyses of 10 million generations were run. Posterior probabilities were plotted and post-burn-in generations were retained and used to generated clade posterior probabilities.

Unless otherwise noted, all parsimony analyses were performed in TNT using a 'New Technology' (NT) search as a basis. Default settings were used with the following exceptions: 'Ratchet', 'Sectorial', 'Fusing' and 'Drift' turned on; 'initial level' set to 50; 'check level' turned on and at default level 3; 'initial addition sequences' set to 10; and 'find minimum length' set to 5. Following the completion of each NT search we performed a traditional search using the trees in memory, holding a maximum of 1000 trees. All gaps were treated as missing data. Results are reported as a strict consensus of these trees.

Results

Method performance

Comparison of results across methods is complex when data are variously included, excluded, or recoded. Because hyper-partitioned data can be modeled in many different ways it may ultimately be possible to use likelihood ratio tests (e.g. Shimodaira and Hasegawa, 1999) to determine whether one topology is significantly better than others. The application of likelihood or Bayesian models to large numbers (hundreds) of small partitions has, however, not been attempted to our knowledge. We use two criteria for judging the methods: 1) congruence among results; and 2) congruence with previously published hypotheses based on morphology (i.e. past taxonomic hypotheses).

Baseline analyses (aligned data with ambiguous regions excluded, Table 5.7, analysis A) present a conservative estimate of the clades that may be present across a range of analysis types. As expected, due to overlapping regions of missing data, and the absence of the "tail" regions (Table 5.5) for a majority of the data, the strict consensus

trees for these analyses were poorly resolved. The consensus for the complete dataset (Fig. 5.4, analysis 13A), does not separate the ingroup (Diapriidae) from the outgroups, but does recover a number of clades that have been previously hypothesized, or that while novel may be reconciled with existing taxonomic hypotheses or morphological data (see 'Implications for Diapriid Phylogeny' below). Excluding the "tail" data (analysis 1A, data not shown), the strict consensus is further resolved, however there are significant problems, including the non-monophyly of the Diapriidae and a high degree of para- or polyphyly. This may be due to the absence of the 18S rRNA data in some of the key outgroup taxa. Bootstrap support values for clades (data not shown) in these analyses were very low, and as such a comparison of bootstrap values across the recoding methods was not undertaken. Bootstrapping of several of the re-coded analyses (data not shown) found similarly low support.

It is held (Lutzoni et al., 2000; Kauff et al., 2003) that ambiguously aligned regions contain phylogenetic signal even though they are frequently excluded from analyses. Hyper-partitioned datasets, wherein these regions are carefully delimited provide an excellent basis for a test of this hypothesis. Combinations of recoded and standard aligned data also reveal increased (e.g. Fig. 5.5) or similar (e.g. 5.6) resolution in some casess, but not all. Analysis of ambiguously aligned data alone illustrates (Figs. 5.7-5.9) that there is signal present in these typically excluded regions. Quantifying this support with partition Bremer support (Baker and DeSalle, 1997; Lambkin et al. 2002) is possible, but these tests are left for future exploration. Both the *k*-word and ARCO recoding methods recovered clades of taxa (Figs. 5.7-5.9, vertically labeled clades and Fig. 5.8, white ovals) that all belong to a single higher taxon. The elision-like approach, including all possible *k*-words for all excluded data, on the rRNA 28S D2 data alone, is remarkably informative (Fig. 5.9). Therein, Ambositrinae - *Pantolytomyia* is recovered as monophyletic, and the poly- and paraphyly of the Diapriinae and Belytinae is greatly reduced. Similar results for elision-like approaches including the 18S and 28S D3-D5 were not recovered, an extreme degree of polyphyly was typically recovered in these results. These contrasting results suggest that recoding data in ambiguously aligned regions should be done with caution.

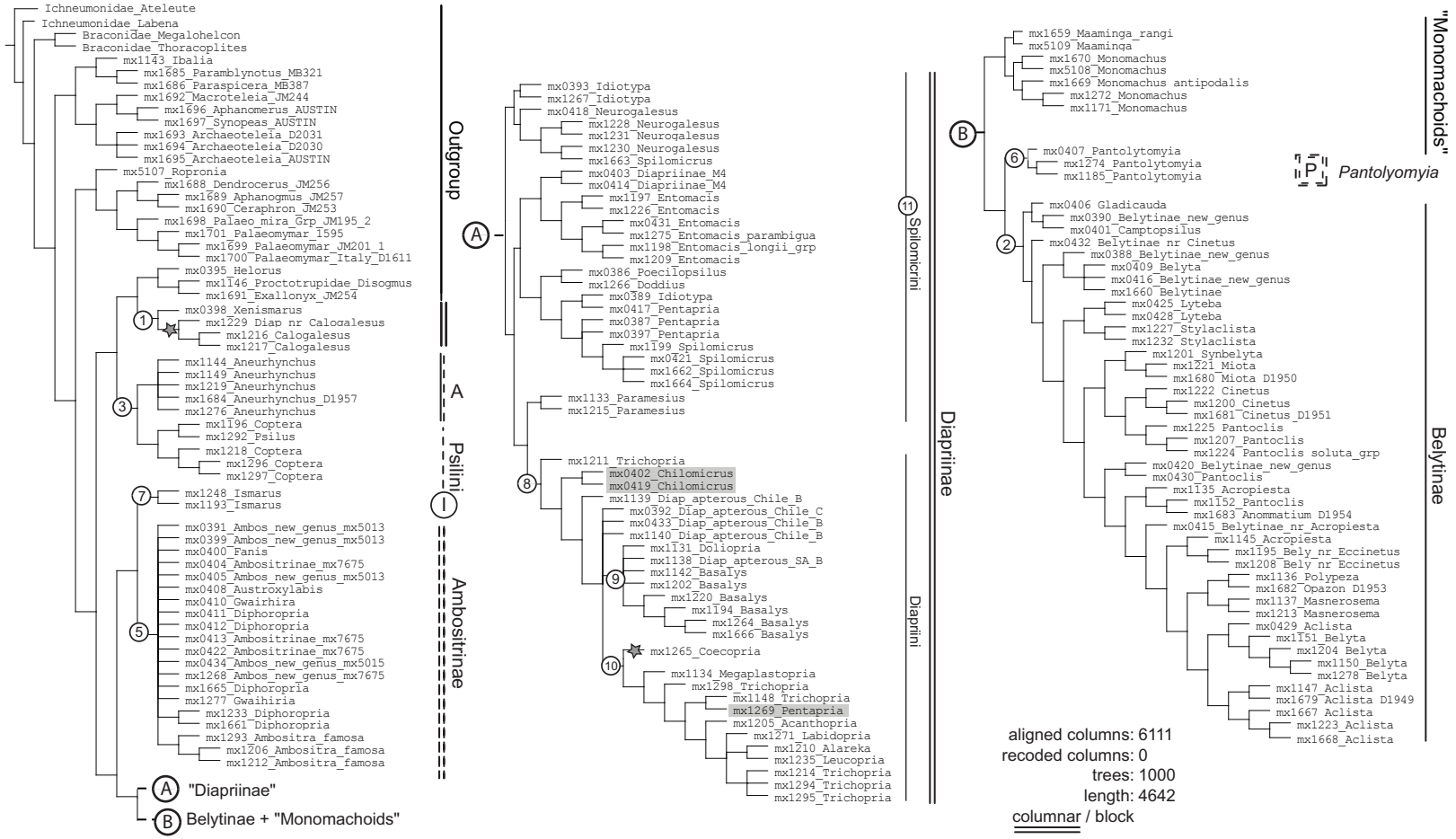

FIGURE 5.4. Results for analysis 13A (all aligned data). The Psilini are presently placed in the Diapriinae. *Pantolytomyia* ("P") is placed in Ambositrinae. *Aneurhynchus* ("A") is variously placed in Diapriinae or Belytinae. *Ismarus* ("I") represents the monobasic Ismarinae. Starred (✯) taxa are considered Diapriinae, incertae cedis. *Xenismarus* and the greyed taxa are placed in the Spilomicrini (Diapriinae). Circled numbers 1-11 are referenced throughout the text.

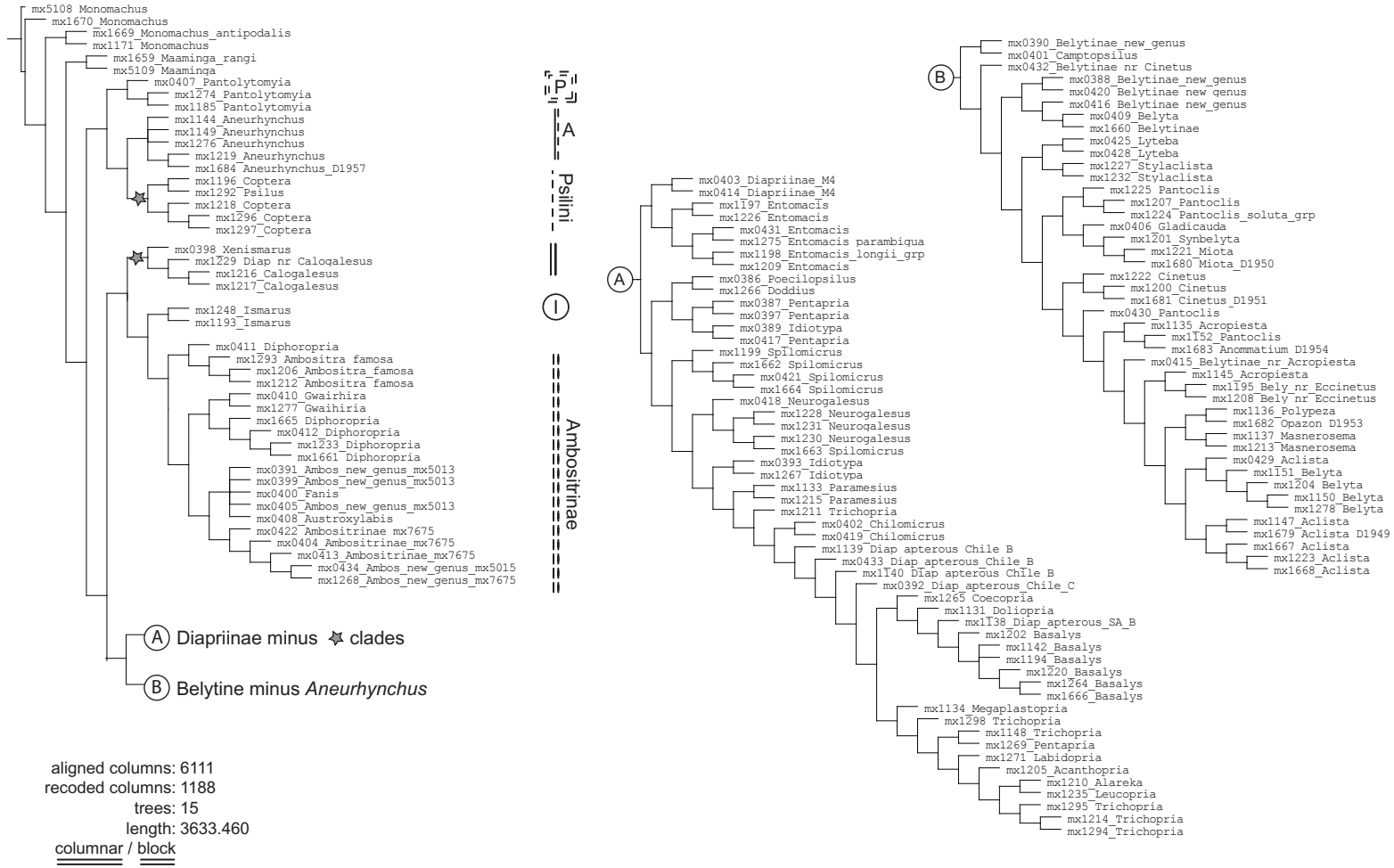

FIGURE 5.5. Results for analysis 14M (ARCO recoding on bracketed partitions with unbracketed partitions included and unmodified).

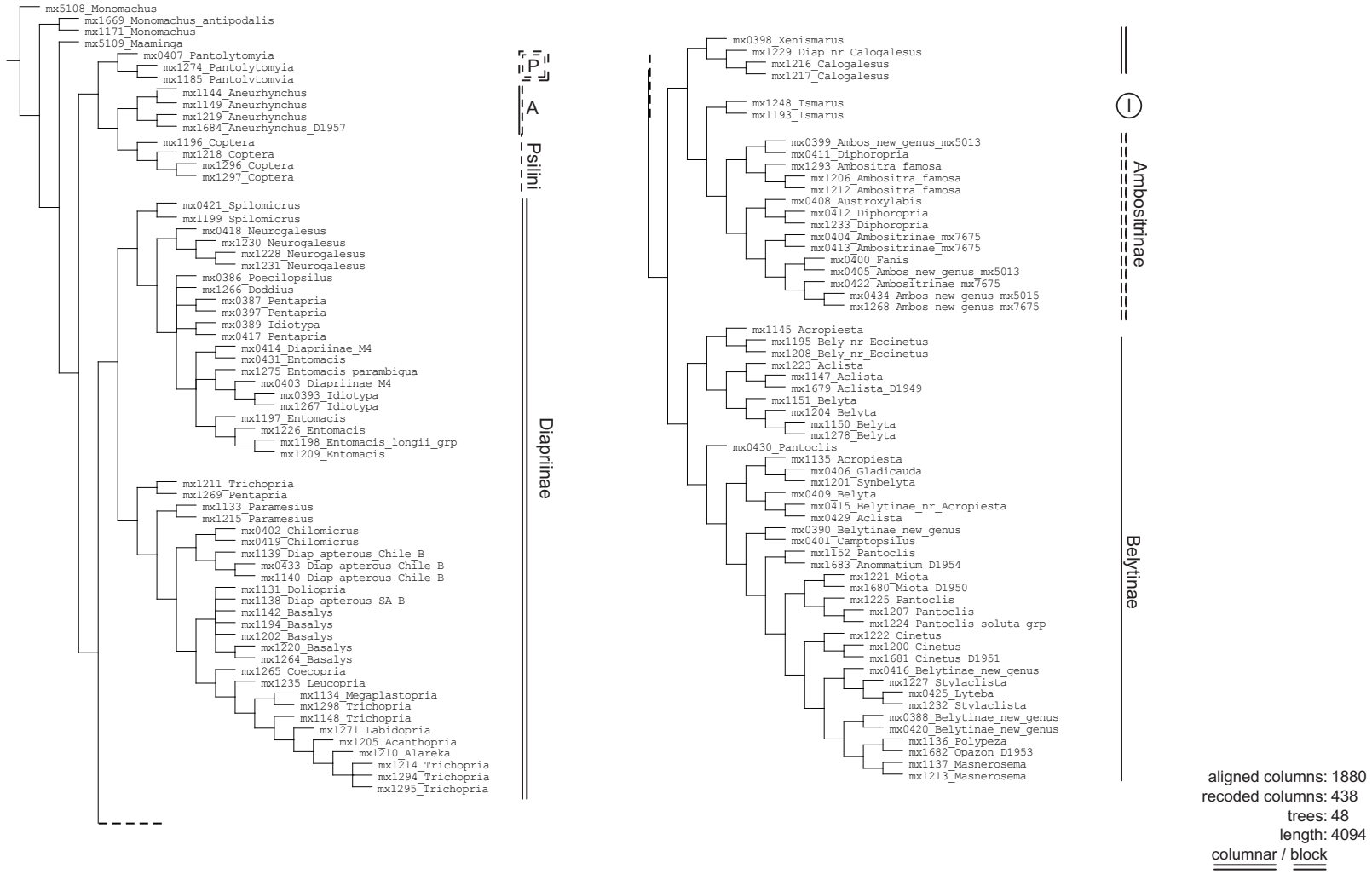

FIGURE 5.6. Results for 2P (28S D2, using *k*-words of length 4).

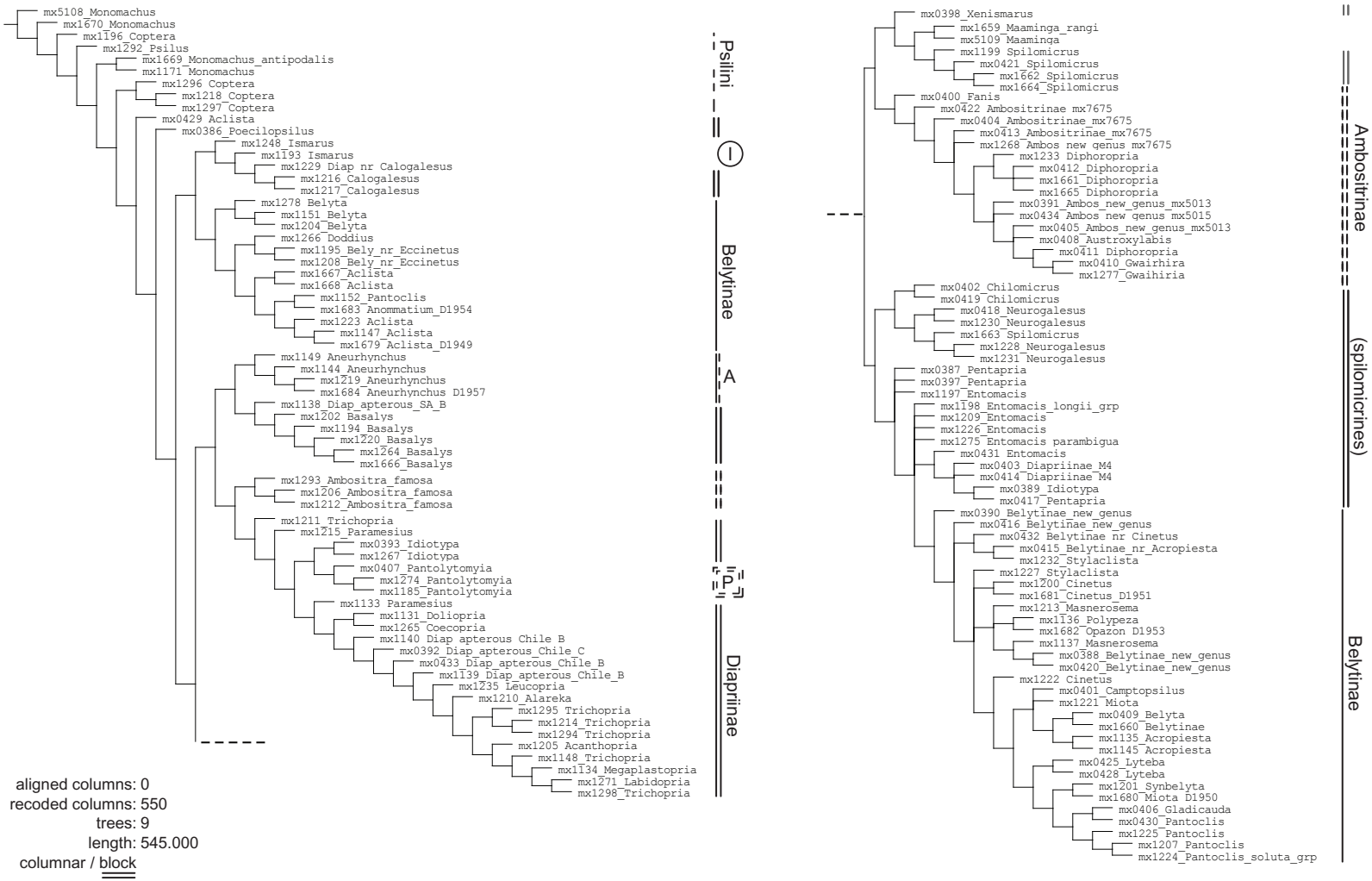

FIGURE 5.7. Results for analysis 6N (28S D2 bracketed data alone and recoded using ARCO characters). Vertical names are not meant to indicate monophyletic groups but rather membership in the given clade.

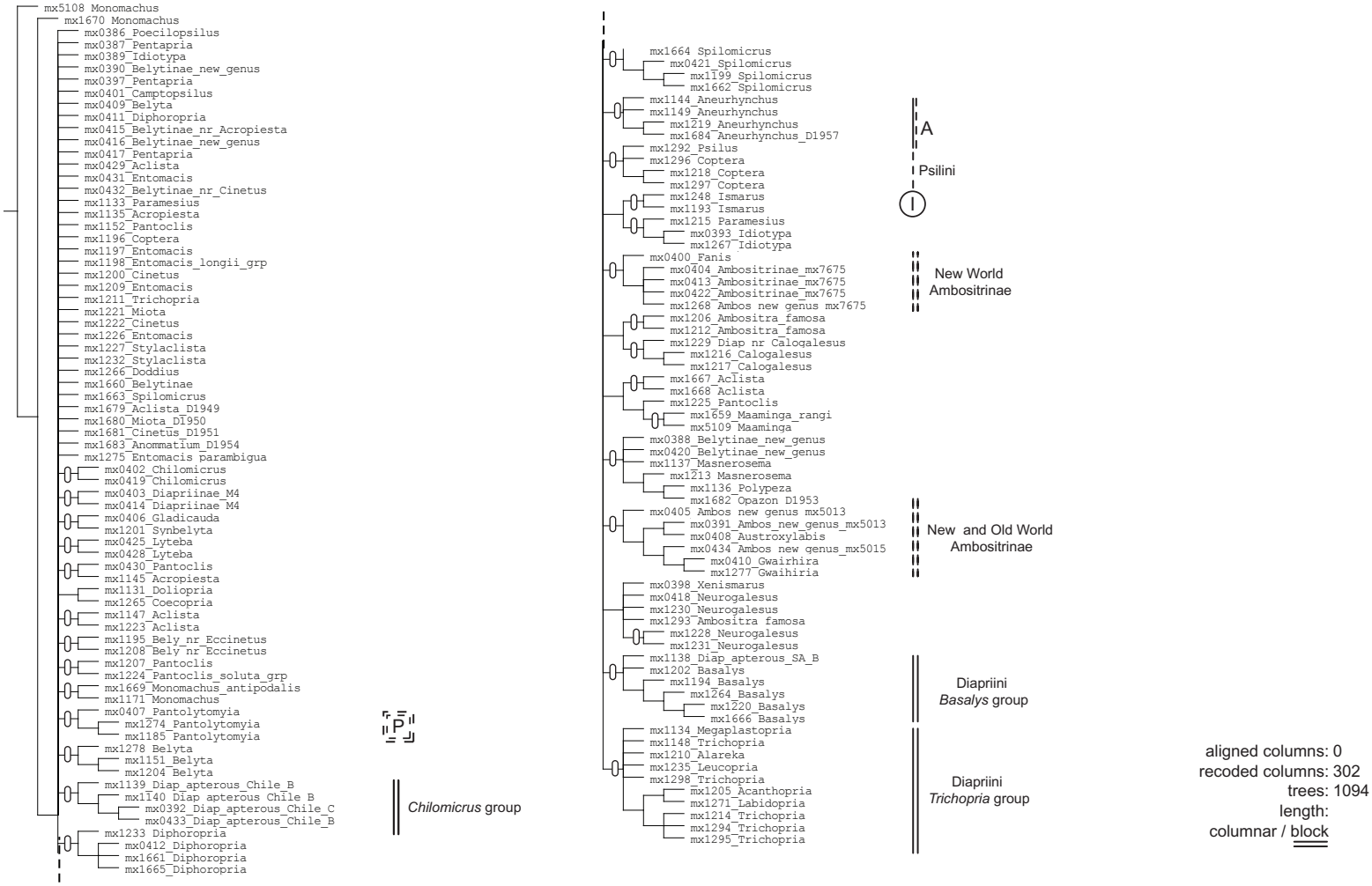

FIGURE 5.8. Results for analysis 6Q (using *k*-word size 4 on 28S D2 bracketed data alone). White ticks mark indicate monotypic groups (i.e. all members of larger monophyletic clades that were recovered in clades in other analyses).

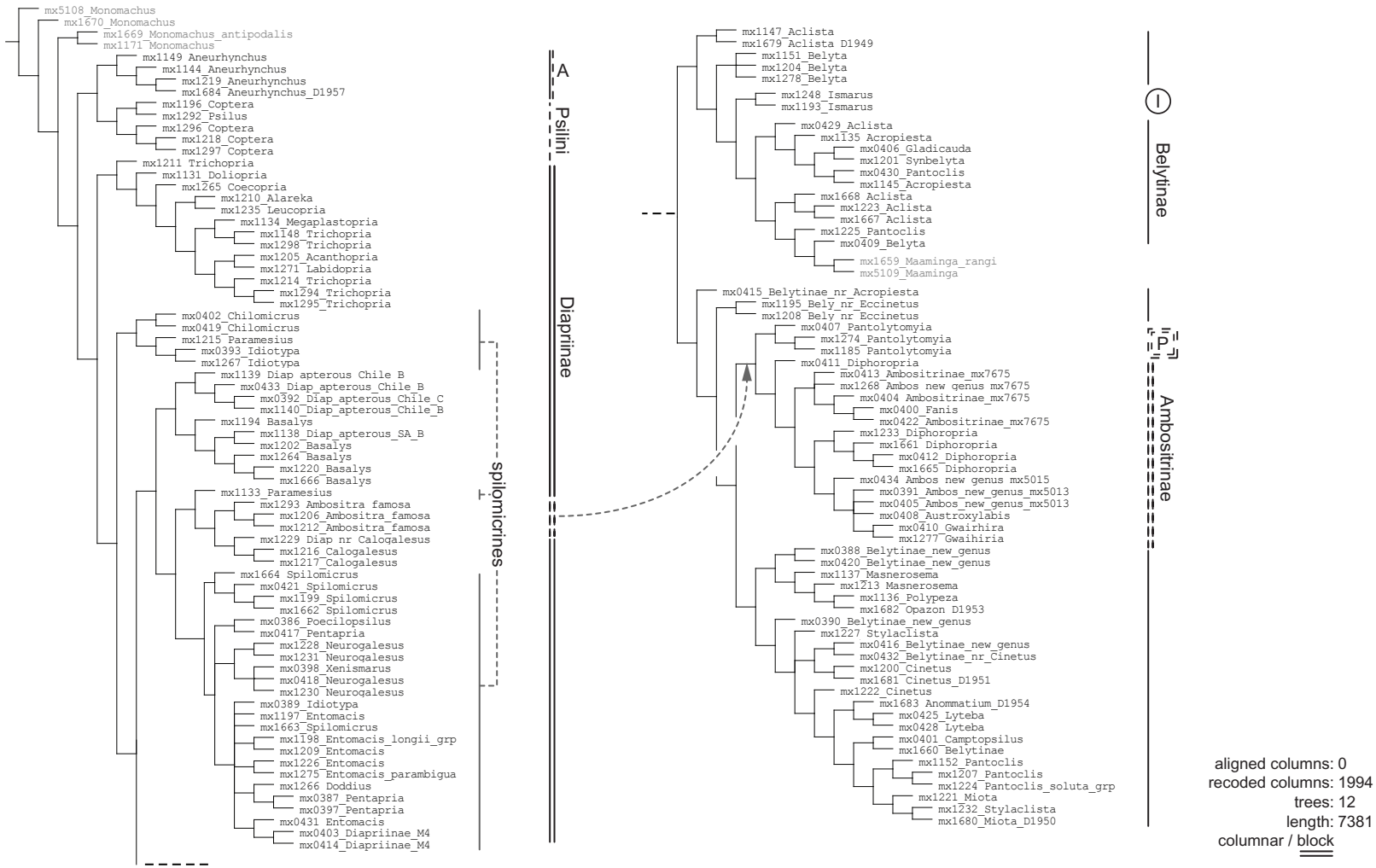

FIGURE 5.9. Results for analysis 6Q (elison *k*-word recoding for all possible *k*-words on excluded 28S D2 alone). Note that *Pantolytomyia* is sister to Ambositrinae minus *Ambositra* (grey arrow).

Given that there is signal recoverable by the recoding of excluded data we may expect that the addition of the recoded data to the aligned data should result in increased resolution in the combined analysis. This result was generally found to be the case (e.g. Figs. 5.5, 5.6), though the increase in resolution did not always equate with consistently recovered novel relationships. For example, the lettered groups indicated in Fig. 5.4 are recovered in the combined analyses, though the relationship of these groups to one another varies (Figs. 5.5, 5.6, and data not shown).

In the case of the combined analyses using the ARCO characters at least some of the increased resolution is likely due to the nature of the ordered multistate characters whose states are real numbers. Optimizing characters whose states are real-numbers results in fewer possible equally parsimonious solutions, and therefore a higher chance of greater resolution (fewer polytomies) in the consensus. The same is not the case for the *k*-word method, though the high number of characters possible in the elision-like case may also be responsible for fewer most parsimonious (MP) solutions. Since scenarios in both recoding cases (*k*-word, ARCO) can be envisioned that will lead to the increase in resolution irregardless of the increase in signal, and since no markedly novel relationships were recovered in the empirical tests provided here, additional empirical tests will be required to explore the effect of adding recoded characters to traditionally aligned data.

Given that the final recoding method proposed here, treating basepaired columns as a single character, reduces by half the number of potentially parsimony informative characters a reduction in resolution is expected. When only the recoded columns are included (e.g. Fig. 5.10) this resolution decrease is noted. The resulting topology is, however, largely congruent with the results of the baseline analysis, and more importantly no potentially spurious relationships are introduced. Including ARCO recoded data (Fig. 5.11) increases the resolution, with the major notable result being the splitting of one subfamily, the Diapriinae, into two major clades roughly following the division between Spilomicrinii and Diapriini, with these clades separately sister to the Belytinae and Ambositrinae respectively. Aside from this result, this analysis is interesting in that all data are recoded based on structural elements, with the resulting

topology agreeing well with morphological data. This suggests that the structural hypothesis coded in the matrix (basepairing hypotheses and partitions) are supported, or at minimum, are not positively misleading.

The recoding of basepairs to a single column retains aspects of the original columnar alignment, and as such is arguably still a traditional columnar-based approach. A more extreme recoding method, recoding all partitions (both aligned and unaligned) was also tested. In this case there are no columnar homology hypotheses, and the analysis can be thought of a test of the structural hypotheses delimited by the hyper-partitioning. Trees derived from this type of analysis are illustrated in Figs. 5.12 and 5.13. With several exceptions, the clades recovered in both these analyses are highly similar to with traditional approaches (Fig. 5.13), however other combinations of data produced less congruent results. Recovering trees based on block-level homology statements only, that are largely congruent with our understanding of relationships derived from morphology, hints that the positional constraints enforced due to hyper-partitioning may indeed represent those that evolutionary forces are selecting on. That is, the partitions defined based on structural elements may be those that are being selected on. If this were not the case (and we make no claim as to having proved this), we would expect to see a higher degree of incongruence in phylogenetic relationships because of the homoplasy introduced by erroneous partitions. While the hypothesis that structural elements identified in hyper-partitioned alignments are those that are being selected on is an exciting one, there are many additional factors that need to be examined to test this. For example, results using the same recoding approaches on the18S rRNA data and the less variable 28S rRNA expansion segments D3-D5 failed to recover similar levels of congruence. It is likely that the degree of signal present is an important factor, i.e. there are ample quantities in the 28S D2 such that a broad spectrum of methods, even given their various flaws, can recover congruent results. More rigorous comparisons (e.g. additional phylogenies generated from other genes) are possible, and should provide further insight in the future.

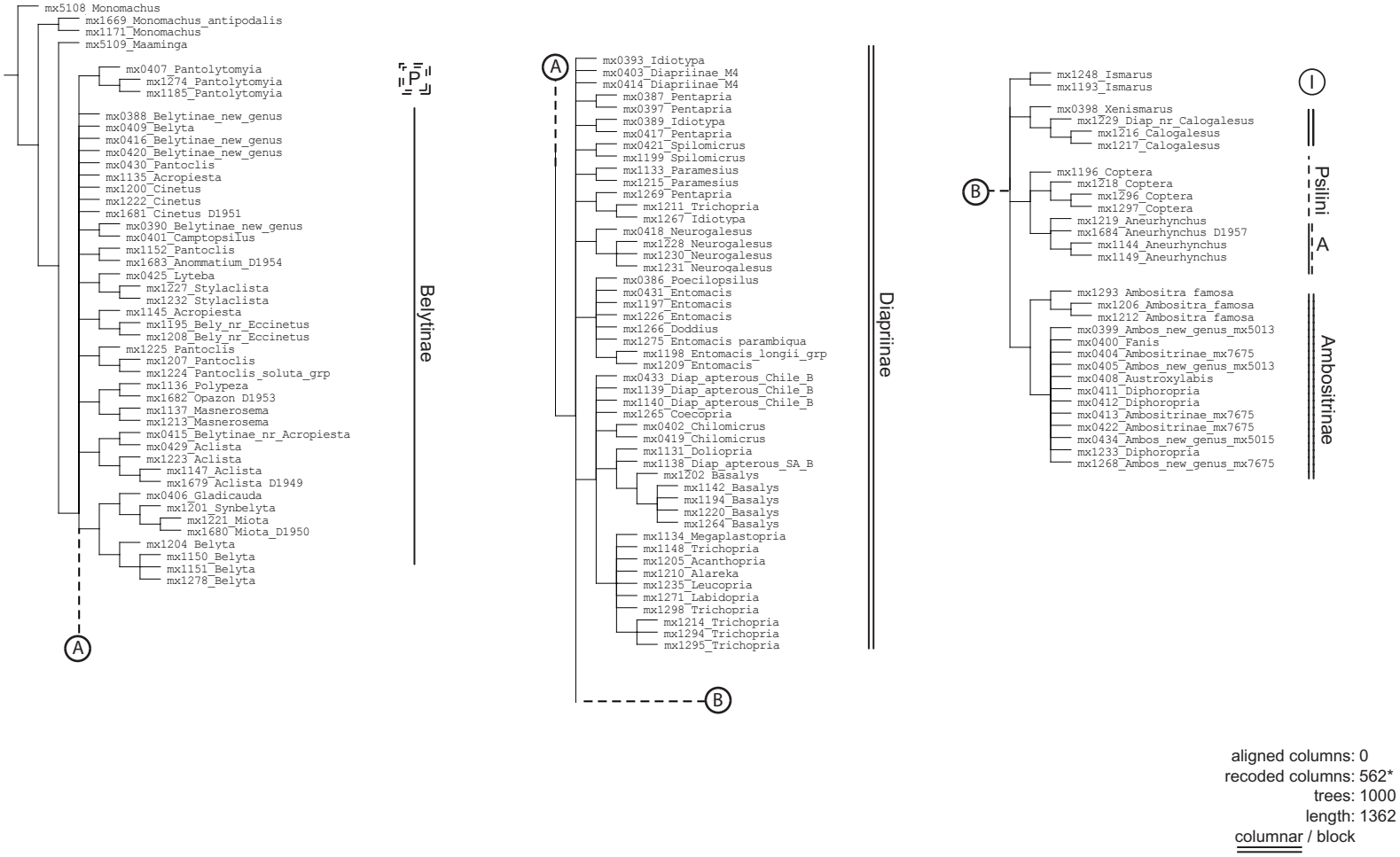

FIGURE 5.10. Results for analysis 2F (recoding of basepairs according to Table 5.3).

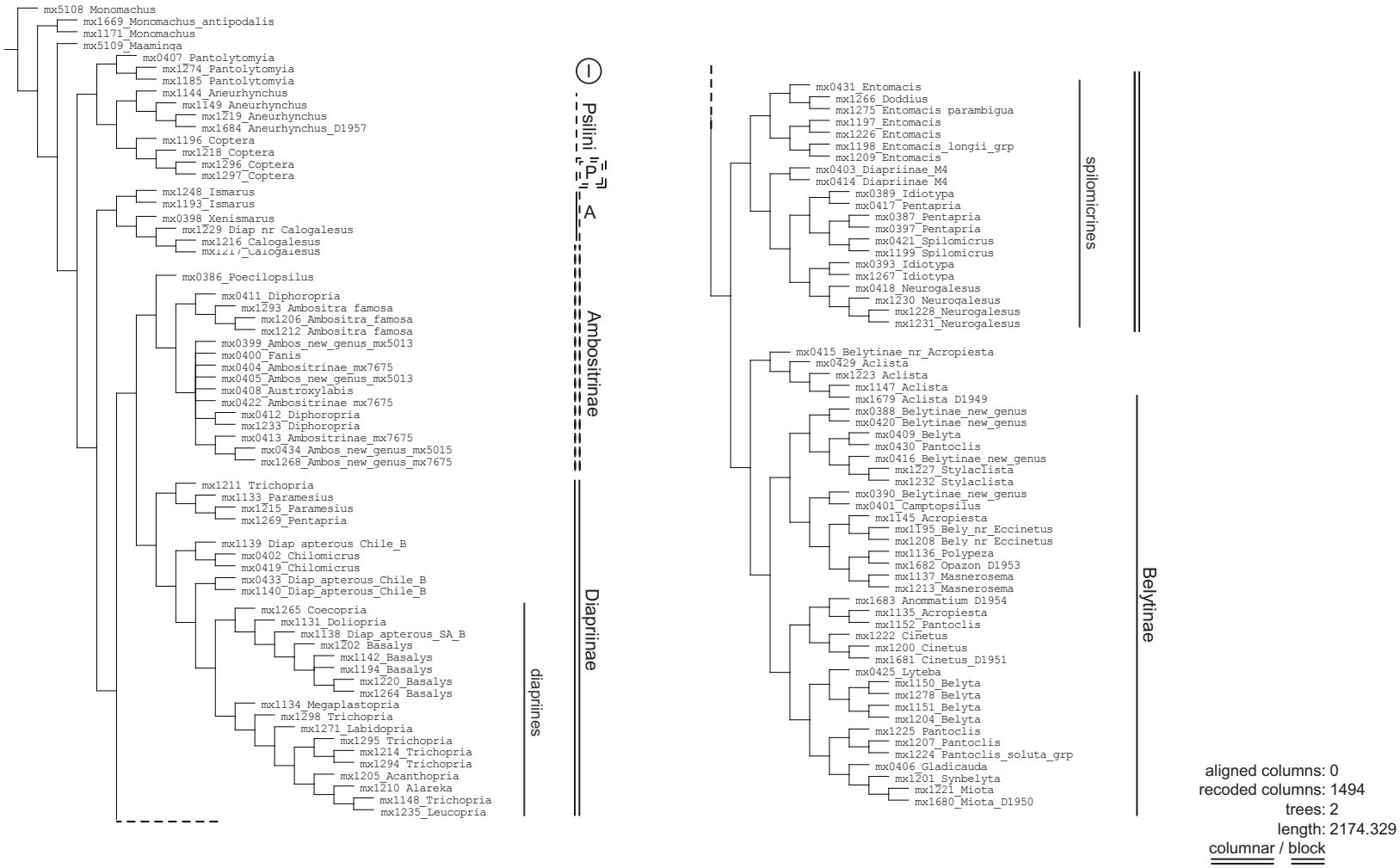

FIGURE 5.11. Results for analysis 2J (28S D2 data, ARCO recoded unaligned data, and pair recoding according to Table 5.3 on aligned data).

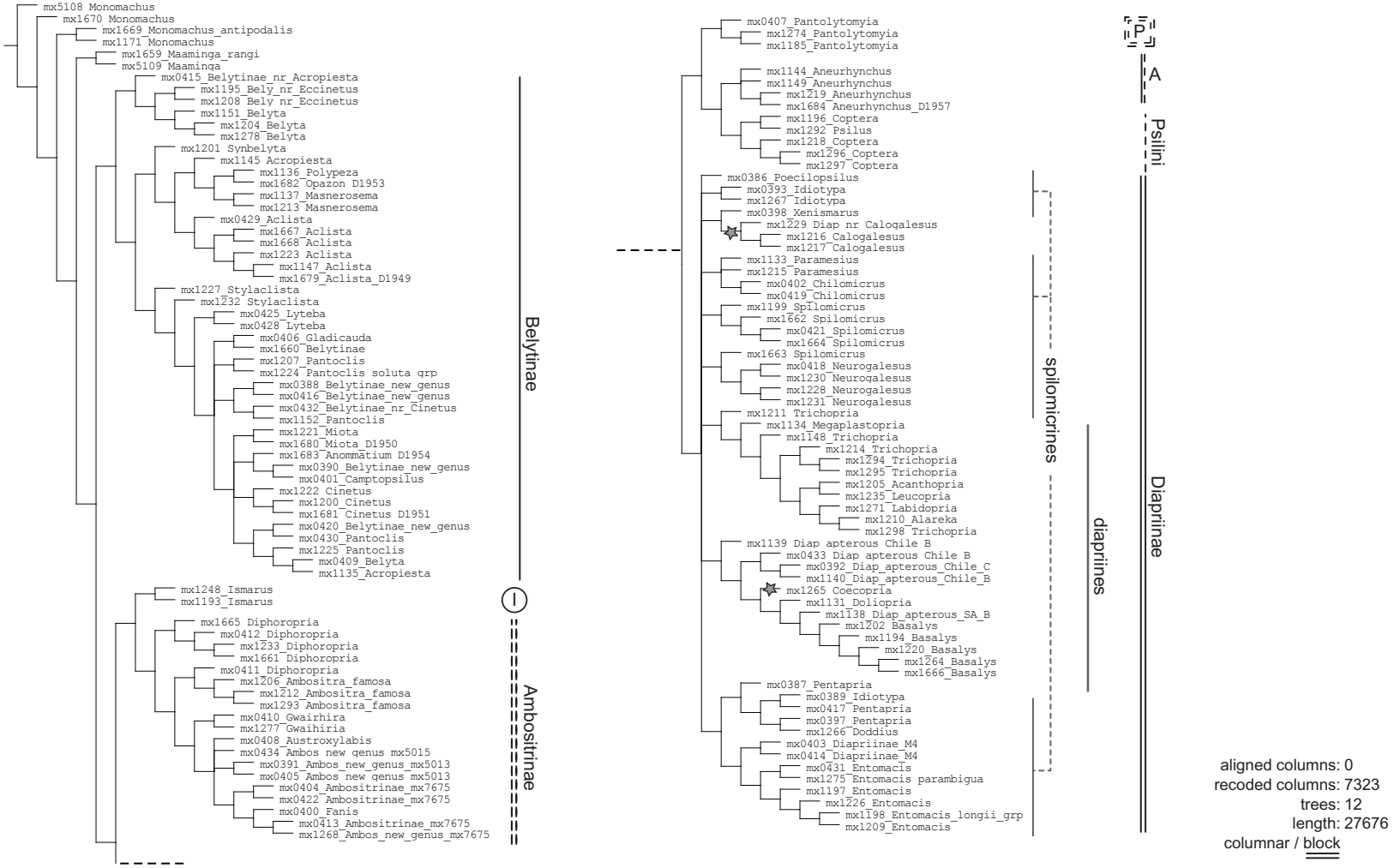

FIGURE 5.12. Results for analysis 6R (28SD2, all data partitions recoded to *k*-words, for all possible *k*-word sizes). Stared clades are Diapriinae, encertae sedis.

FIGURE 5.13. Results for analysis 4O (all partitions recoded as ARCO characters, i.e. there are no columnar homology statements present in the analysis).

Given the relative initial success of the *k*-word method we sought to further explore its usefulness and implications as a more generally applicable approach. We initially envisioned the *k*-word coding method as a rapid way to translate, in an unbiased manner, the data in unaligned partitions into informative characters. Recognition that partitions may in fact interrupt motifs that were created from an informative mutational event lead us to explore the effect of fusing partitions. Subsequent exploration of the method on larger partitions (fusions of smaller partitions) provided results that suggested that the method was applicable to such sized datasets. An overview of the information content in these matrices can be seen in Figs. 5.14-5.20. Figures 5.14-5.17 plot the number of parsimony informative *k*-words by partition. Variation in the number of small (e.g. 3-8) *k*-words present (Fig. 5.14) reflects the rate of evolution of a given partition, as the number of small words increases with both the number of mutational events and the increased density of these events. Note that the theoretical total possible number of *k*-words possible ($4^{\text{word size}}$) is much higher than realized in the actual data. It may be possible to use *k*-word plots as a means to comparatively examine gene heterogeneity. Figures 5.15 and 5.16 illustrate the number of words per delimited variable region for each of the 18S and 28S rRNA data respectively. Note that some of the variation illustrated is due to the number of taxa present, for example expansion segment D3 (Fig. 5.16) is sequenced for many more taxa than the remaining segments. The number of parsimony informative *k*-words may also be used as a means of looking at the relative differentiation between two groups of taxa (Fig. 5.17, "A"). In this case plots of combined groups (e.g. Fig. 5.17, Ambositrinae) can be compared to plots of that group with a set of taxa removed (Fig. 5.17, Ambositrinae-*Pantolytomyia*). In theory, proportionately smaller decreases in the number of words present should reflect proportionately higher degrees of relatedness.

K-word data can also be visualized by plotting the actual matrix of presence/absence characters (Figs. 5.18-5.20). These plots quickly illustrate the relative degree of conservation in the data (e.g. compare 18S vs. 28S D2 data in Fig. 5.18). For a given column, horizontal bands in Fig. 5.18 are *k*-words or groups of similar words that are shared across a majority of taxa. When taxa are arranged in "phylogenetic" order

within a matrix, shared states (e.g. Fig. 5.19, "B") become evident, frequently as offset bars. Additional patterns are evident in sequentially incremented plots of *k*-word size (Fig. 5.20, ellipse). Patterns that repeat at different scales are termed fractal. In theory, the elision-like approach to analyzing these data (e.g. including all data in Figure 5.20 in a single analysis) should amplify the signal from these fractal-like patterns, while decreasing the signal from noisier, pattern-free areas.

An example of the effectiveness of the *k*-word approach on fused partitions (i.e. unaligned data) is illustrated in Figs. 5.21-5.23. The elision-like approach (Fig. 5.21) on the 28S D2 data alone recovers nearly all the clades identified in the baseline analyses (Figure 5.4, small letters). Similar levels of resolution were not recovered in the analysis of data for the other individual amplicons, however in those analyses there were typically few spurious relationships implied. The probability of *k*-words occurring because of chance alone decreases with increasing motif length. Longer shared motifs are thus potentially better candidates as parsimony informative characters. Analysis of the same data from Fig. 5.21 for word size 8 is illustrated in Fig. 5.22. In comparison of the two approaches a higher degree of congruence is found in the single word approach (e.g. note the monophyly of Belytinae in Fig. 5.22). Note that only words of size 3-15 are used in the elision-like approach illustrated in Fig. 5.21. The topology may change if additional word sizes are included, however changes in topology derived from these cases would likely indicate extremely weakly supported nodes. One unbiased way to determine which word sizes to include would be to use all informative words up to the maximum number/word size, i.e. the point indicated in Fig. 5.14. The *k*-word approach appears to be relatively robust to more variable data. Figure 5.23 illustrates the analysis 1U for word size 8, or all data for all taxa. Most of the spuriously placed taxa were missing data (Fig. 5.23, stars) for one or more of the amplicons. This analysis was one of the few that placed *Xenismarus* and *Paramesius* in the Spilomicrini, a result congruent with morphological hypotheses.

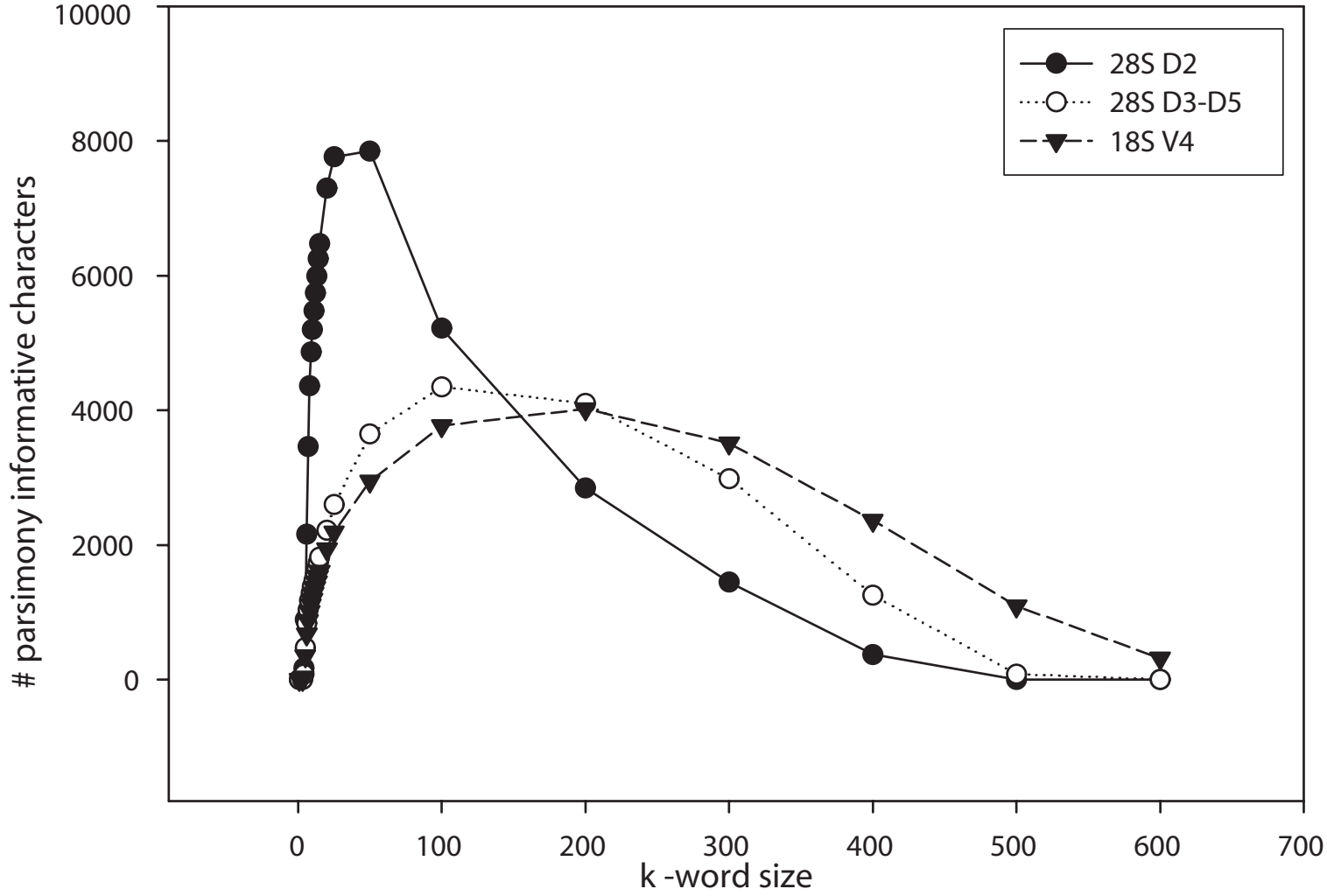

FIGURE 5.14. Plot of the number of parsimony informative characters by *k*-word size for each of the three amplicons examined in this study.

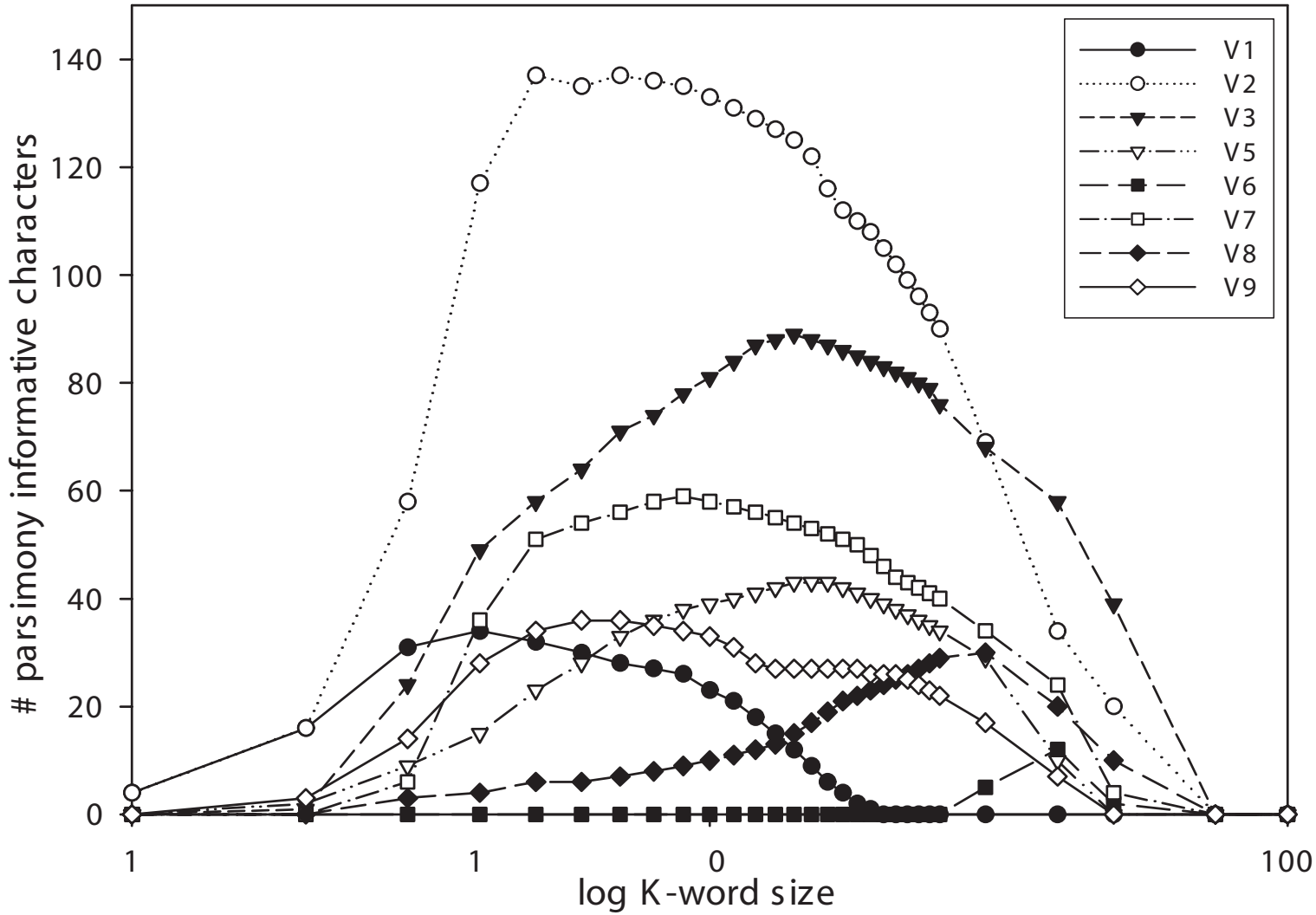

FIGURE 5.15. Parsimony informative characters for expansion rRNA 18S expansion segments.

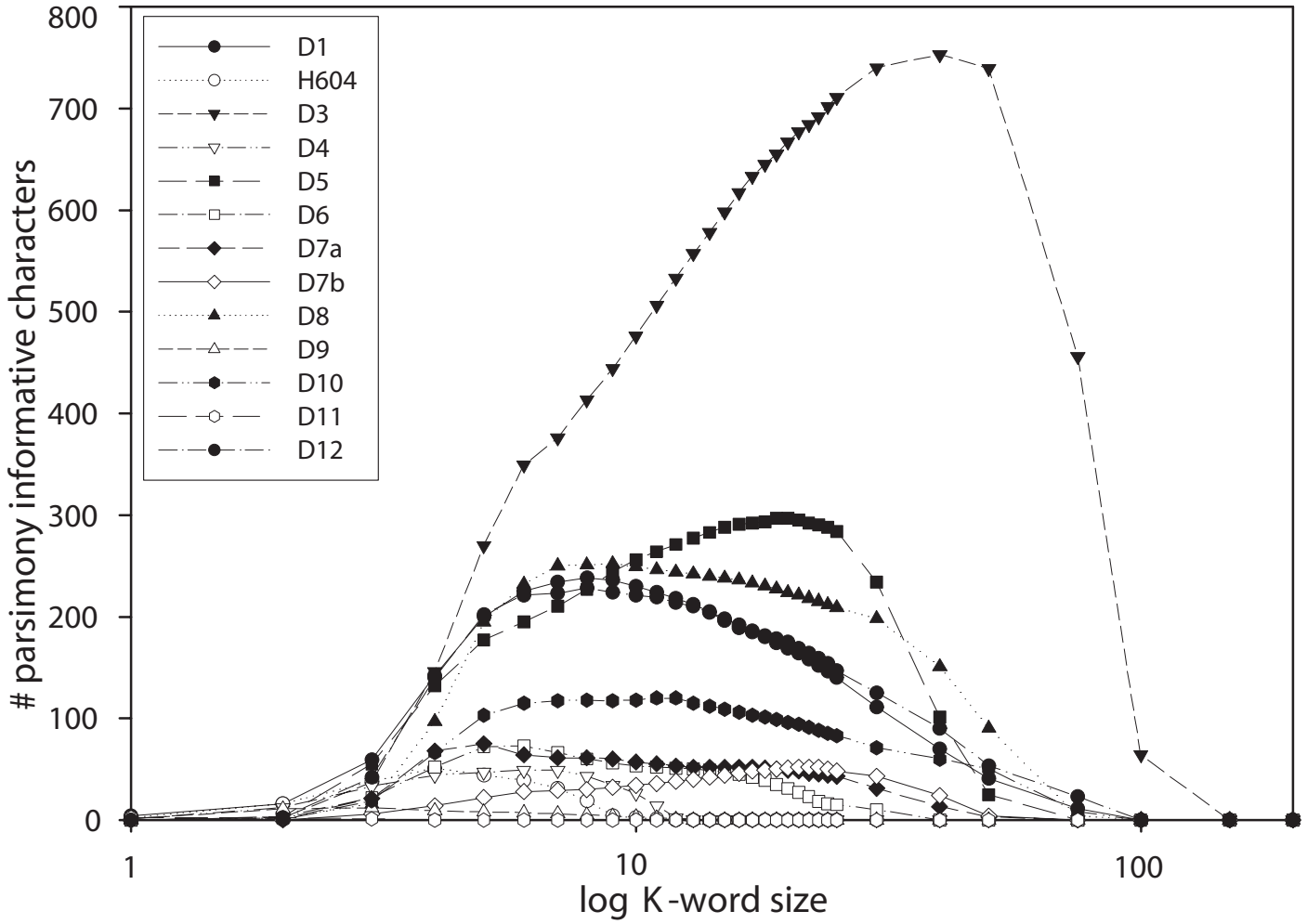

FIGURE 5.16. Parsimony informative characters for expansion rRNA 28S expansion segments. The D2 region is excluded.

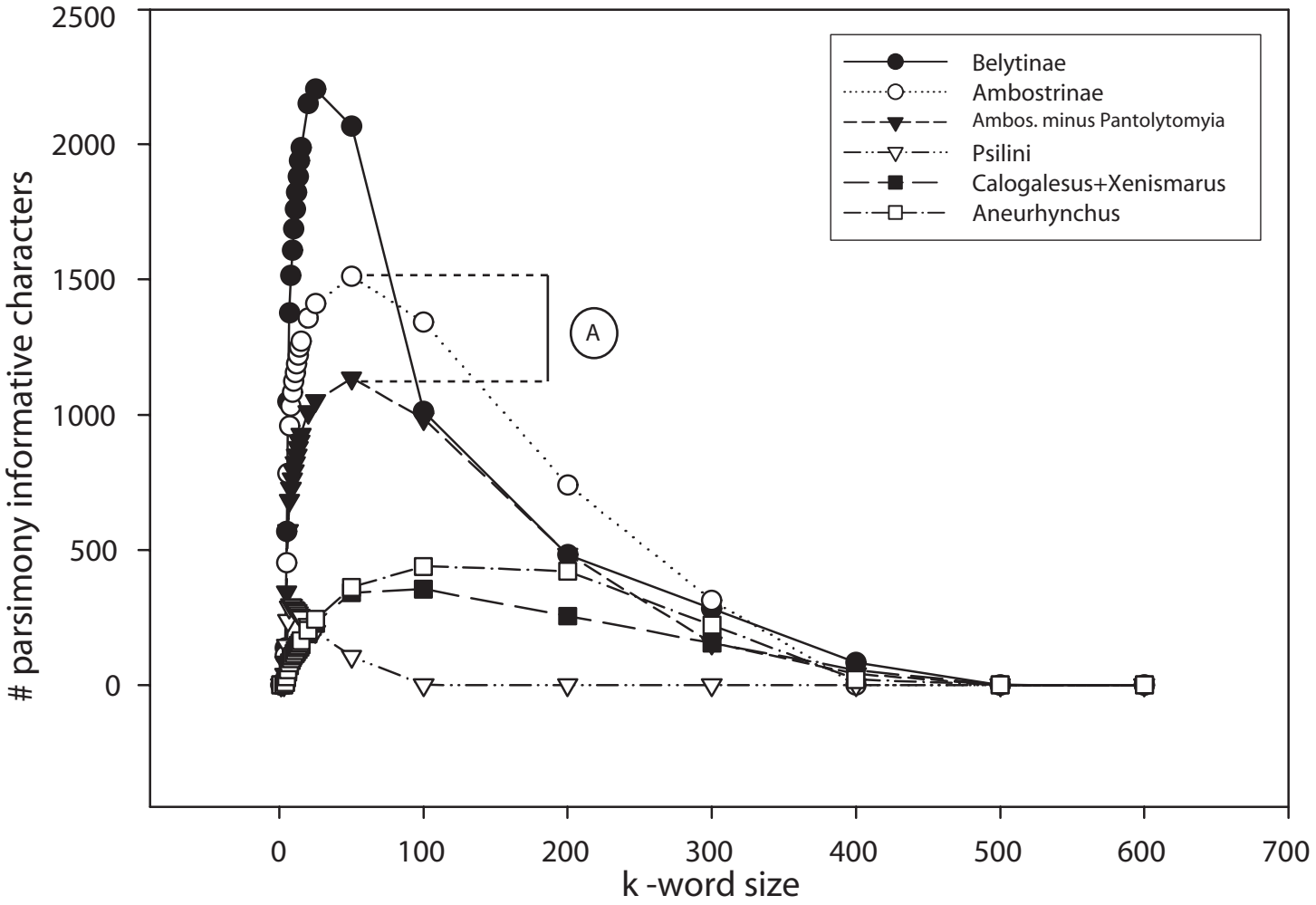

FIGURE 5.17. Parsimony informative characters for several exemplar clades (rRNA 28S data). Relative distances (e.g. "A") between peaks can be used as a proxy measure for the degree of differentiation of one group to another. See text for discussion.

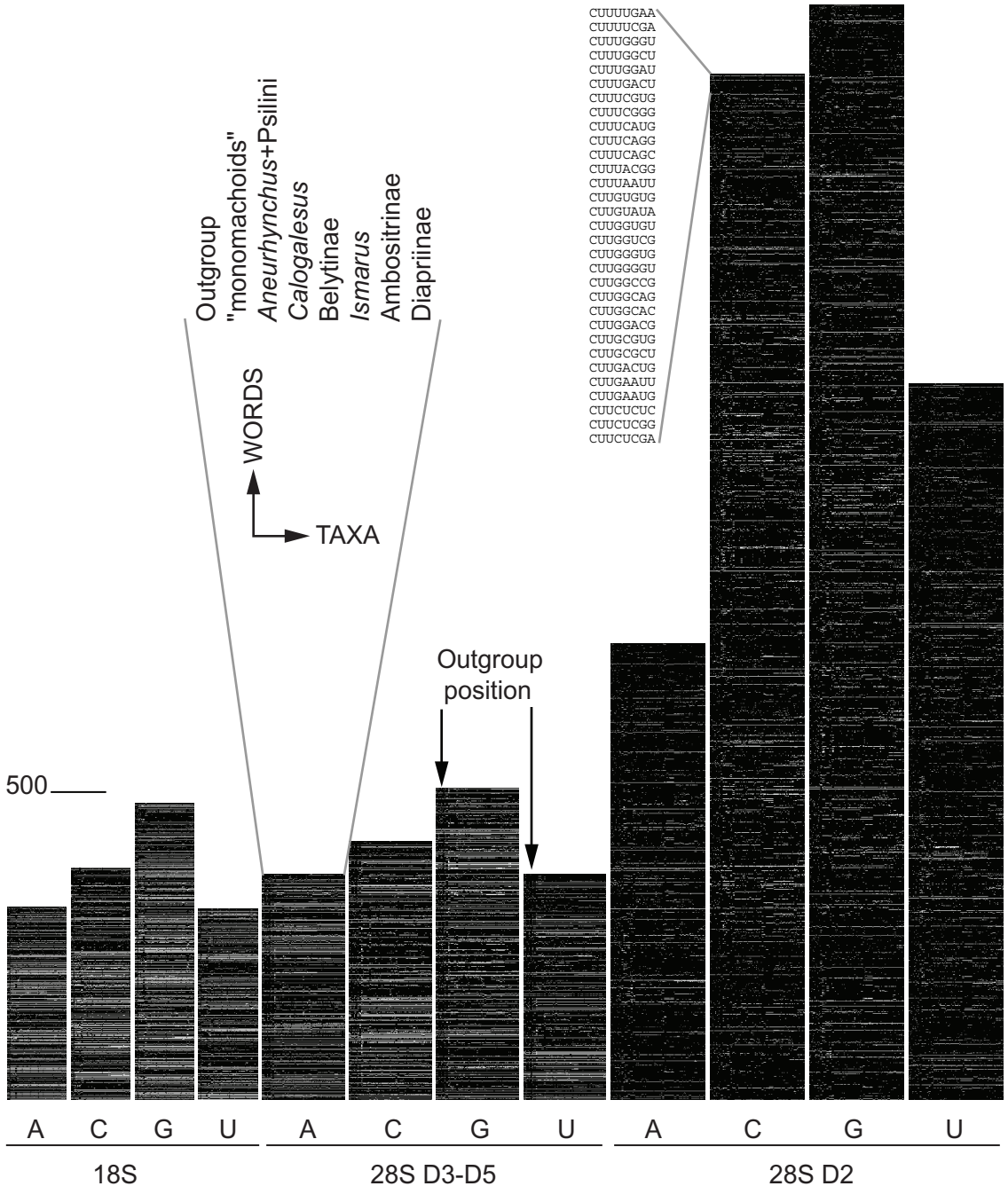

FIGURE 5.18. Variability among the three amplicons in number of parsimony informative *k*-words of size 8. Presence of a given word is given by a grey or white point, absence by black. Brighter shades represent higher GC content for the given word. Within a given bar (e.g. 18S 'A') words are sorted alphabetically from bottom to top. Note that the more conserved gene, 18S, has more consistent bands across the whole matrix. Bar height reflects the total percentage of each nucleotide present.

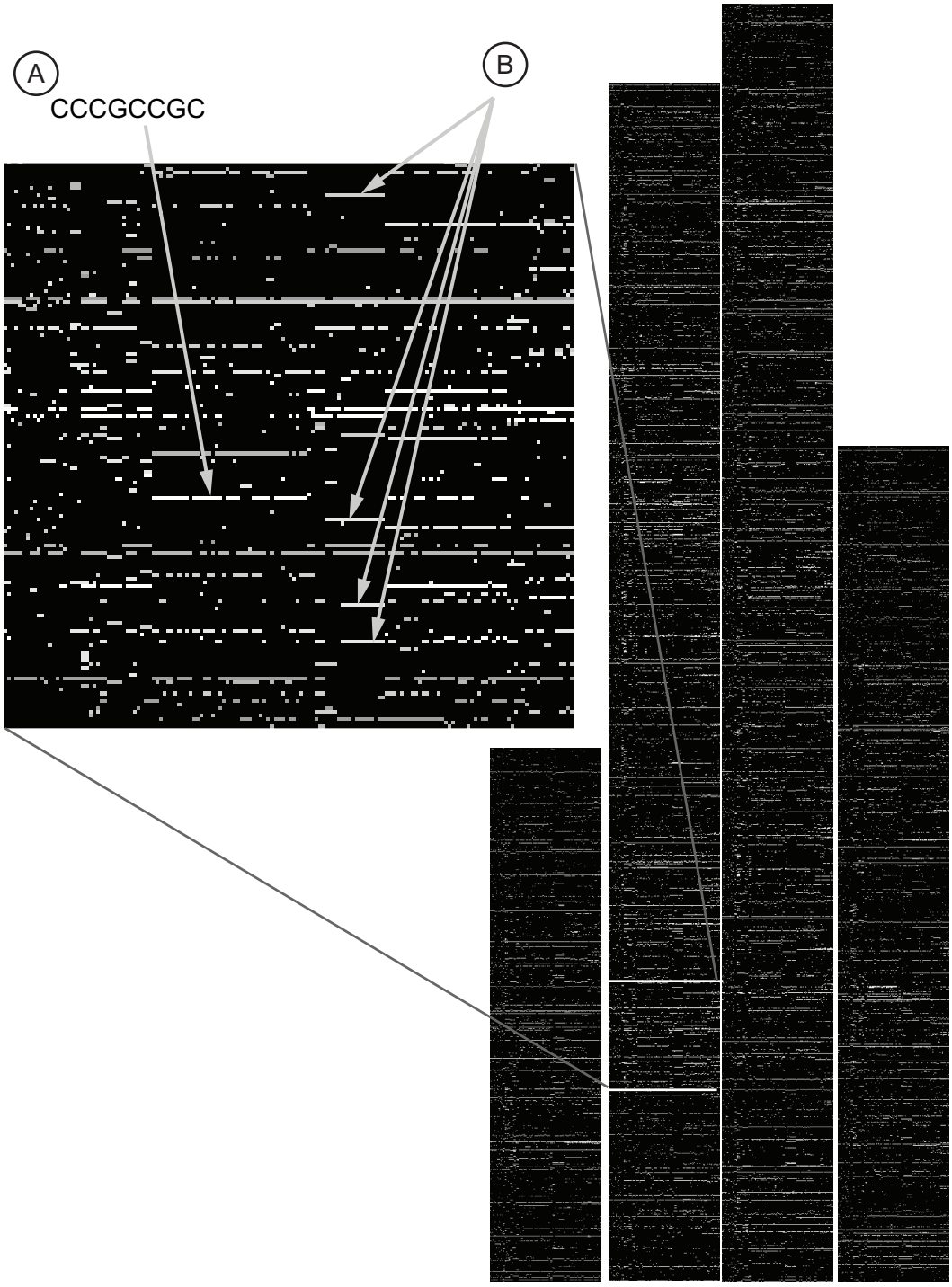

FIGURE 5.19. 28S d2 mapped as parsimony-informative *k*-words of size 8. Columns from left to right are words starting with A,C,G, and U. Brighter shades indicate higher percentage of GC, "A" indicates the presence of word "CCCGCCGC", primarily occuring in the Belytinae and Diapriinae. Offset bands can indicate characters shared by a group of taxa, e.g. "B" highlights bands shared by ambositrines.

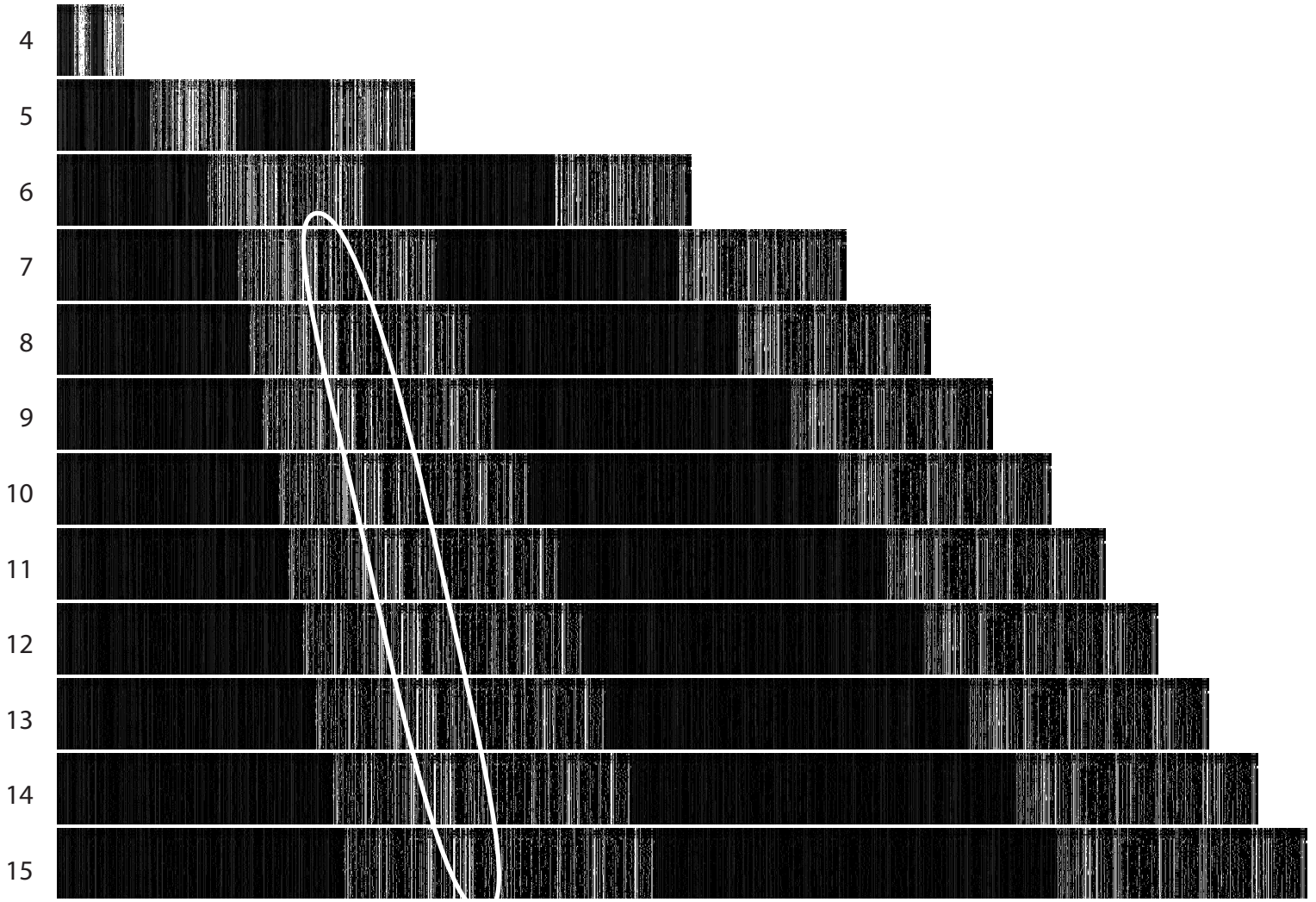

FIGURE 5.20. 28S D3-D5 *k*-word plots. Alternating bands from left to right indicate words begining with A,C,G, and U. Ellipse highights repeating fractal-like patterns that repeat at each increment.

We tested the *k*-word approach on an additional dataset. A reanalysis of Deans et al.'s (2006) unaligned 28S rRNA data alone for *k*-word size 8 recovered every major node (Fig. 5.27, mapped posterior probabilities) that they recovered using additional data and a Bayesian approach (Deans et al. 2006, Fig. 5.6).

The success of the *k*-word approach on unaligned (i.e. non-partitioned) data is not necessarily support for an argument against the hyper-partitioning method, however, it can be used to illustrate potential problems with the approach. In Figure 5.22 all unambiguous character state transitions were found for the node supporting a monophyletic Belytinae, a result recovered in many of analyses. Of those 40 transitions (Fig. 5.22, boxed strings) only 8 were identifiable in the original partitioned matrix, as contiguous strings, and all of these were found in large expansions unique to outgroup taxa. That no single word was recovered as contiguous in the ingroup suggests that the partitioning of the original alignment (i.e. hypothesis) is incorrect in places.

Potential errors in hyper-partitioned alignments can also be identified using traditionally aligned data. Figures 5.24 and 5.25 illustrate a traditional algorithmic approach, using the multiple-sequence aligner MAFFT (Katoh et al. 2002, 2005). The first analysis (Fig. 5.24) was the only one performed that recovered a monophyletic Ambositrinae, a very well supported clade based on morphology. Using a comparative approach, identifying the precise location in the molecule where the state transitions supporting this monophyly occur, may help to identify erroneous partitions in the hyper-partitioned matrix.

The hyper-partitioned framework, and various translations available to it, lend itself to a wide range of additional analysis types. One such approach, treating *k*-word characters as morphological states under Bayesian inference is illustrated in Figure 5.26. The approach returns a remarkably high number of highly supported clades (Fig. 5.26, clades marked with 1.0), at least some of which appear spurious. This analysis was attempted as a preliminary experiment only; however, even in its rudimentary form it illustrates a high degree of congruence with traditional methods.

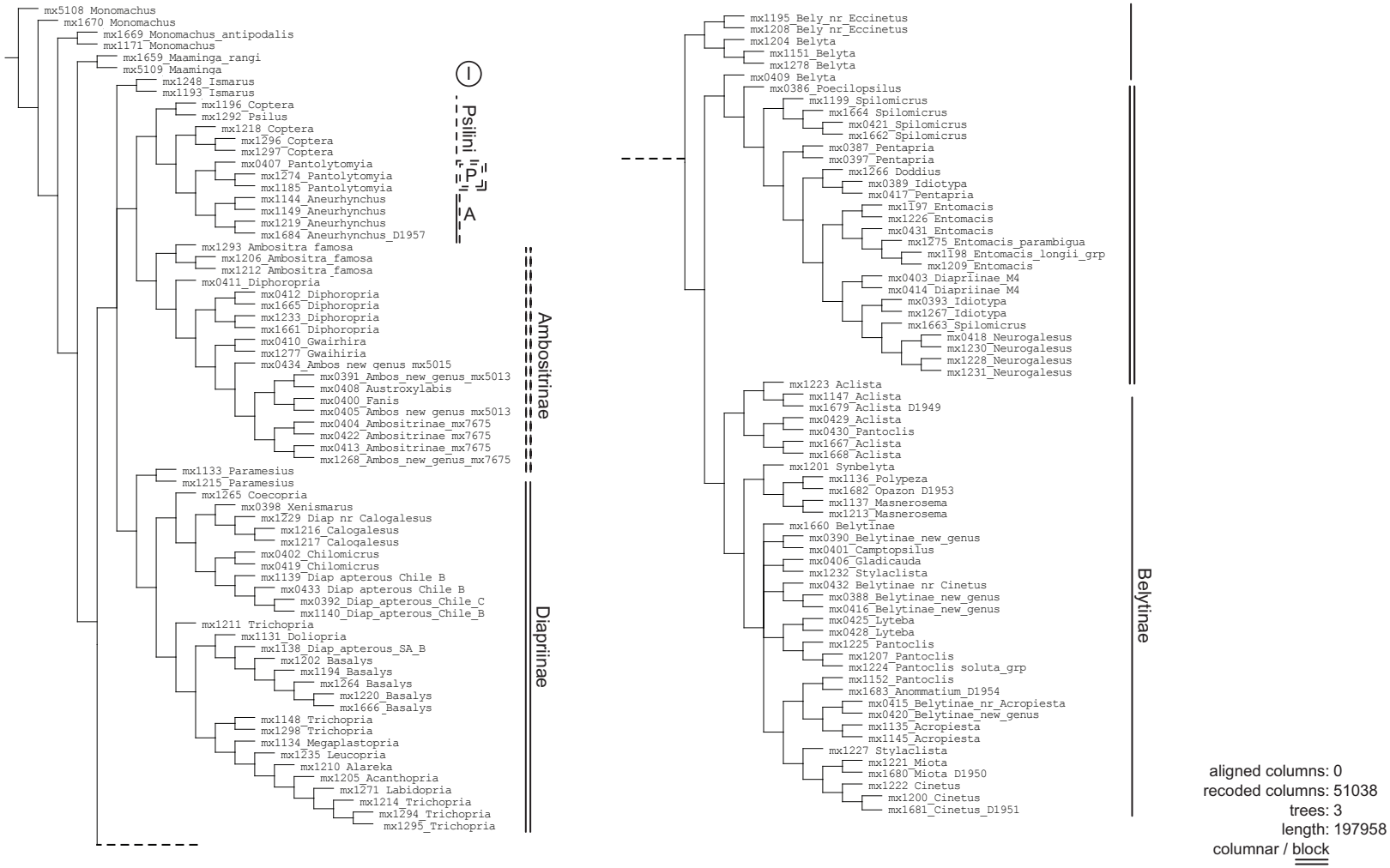

FIGURE 5.21. Results for analysis 6U (k-words size 3-15 on all partitions of 28S D2 data). The tree is derived from completely unaligned data.

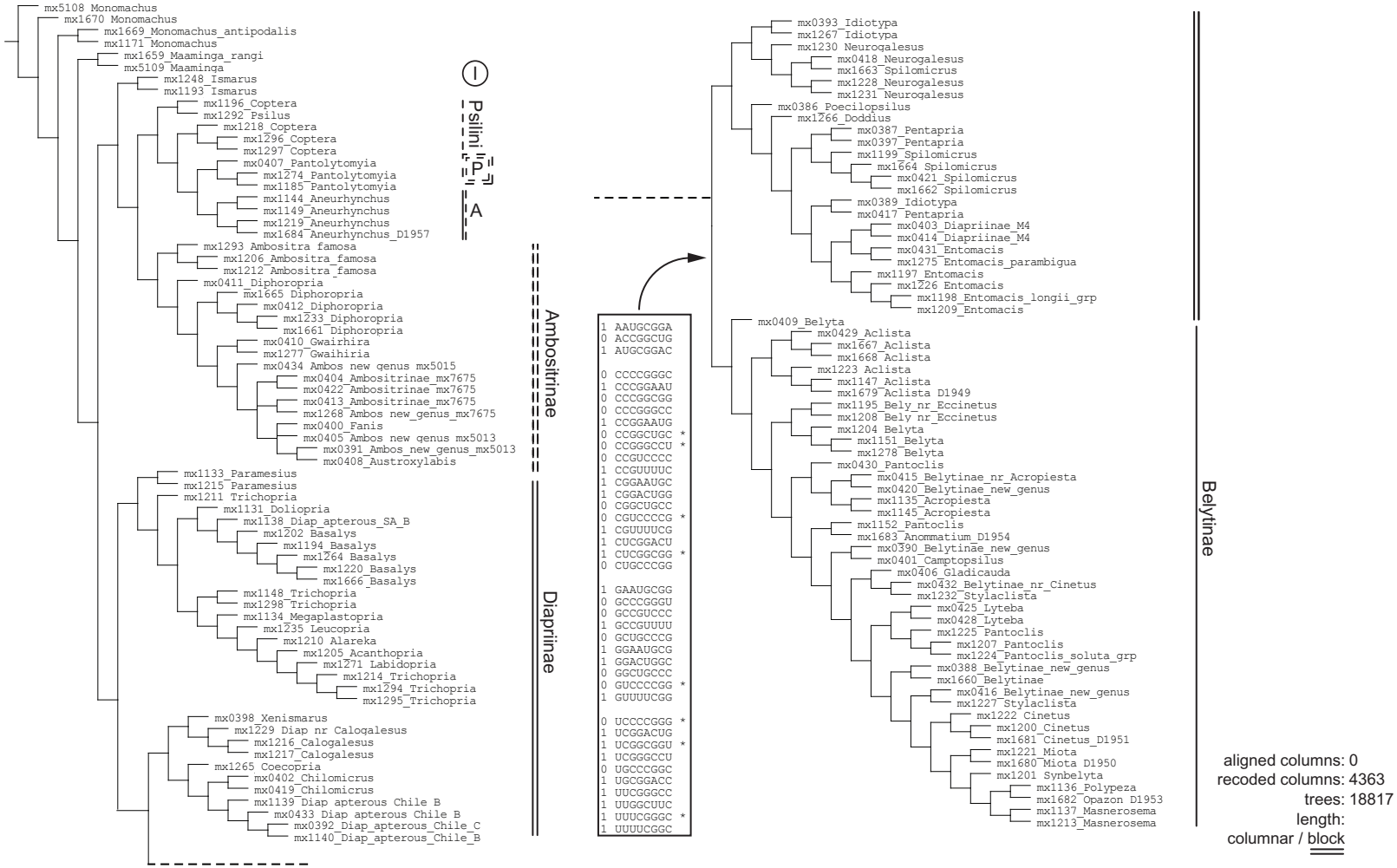

FIGURE 5.22. Results for analysis 6U (*k*-word size 8 on unaligned 28S D2). Boxed data is list of all unambigous state changes (presence: 1, absence: 0) of the given character words at the node which defines Belytinae; those words that are starred were contiguous in the partitioned alignment. See text for discussion.

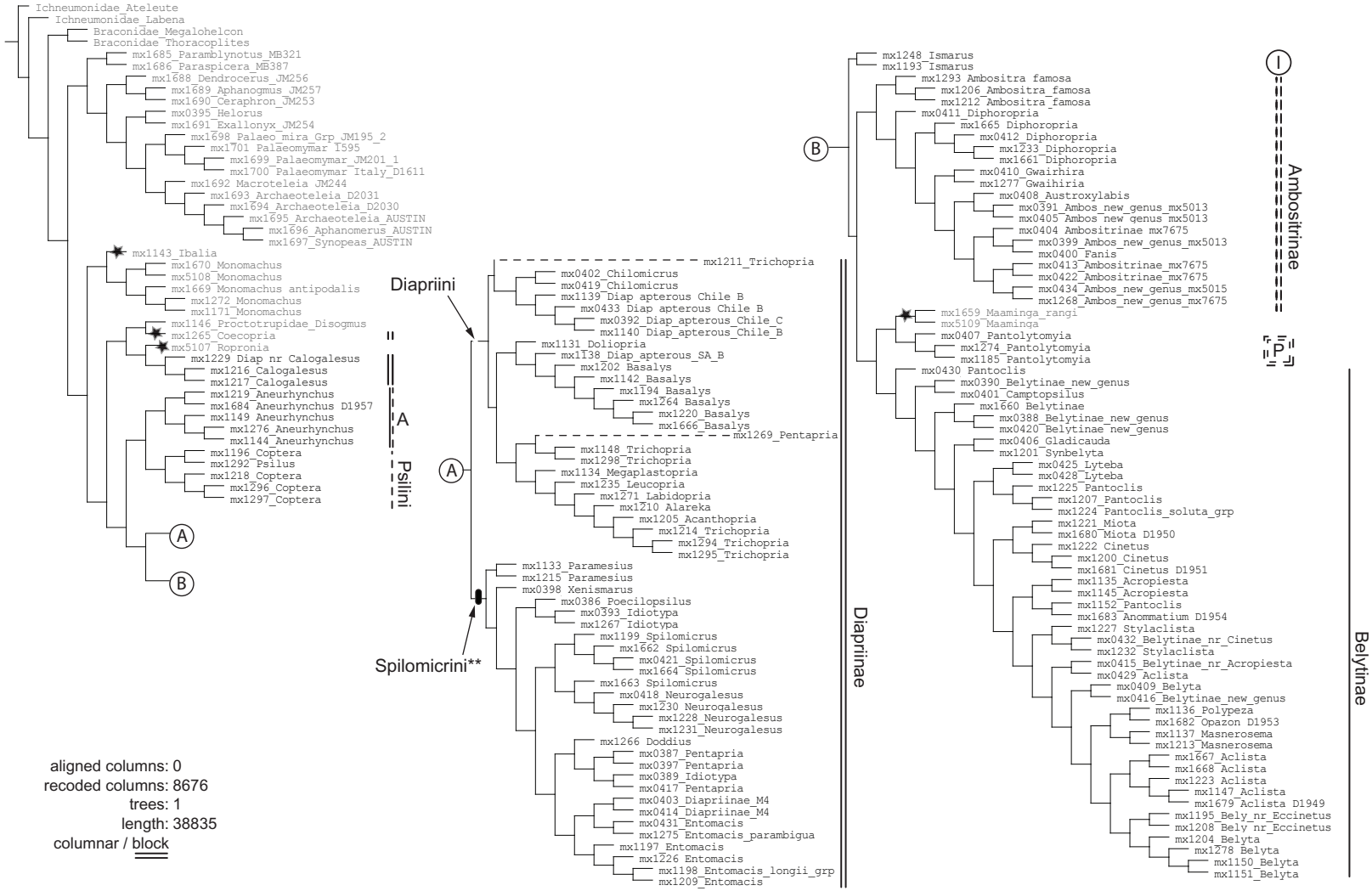

FIGURE 5.23. Results for analysis 1U (all data unaligned, and recoded at *k*-word length 8). Starred (✶) clades/taxa were missing the 18S amplicon. ** Only mx1269_Pentapria and species of *Chilomicrus* are misplaced from Spilomicrini.

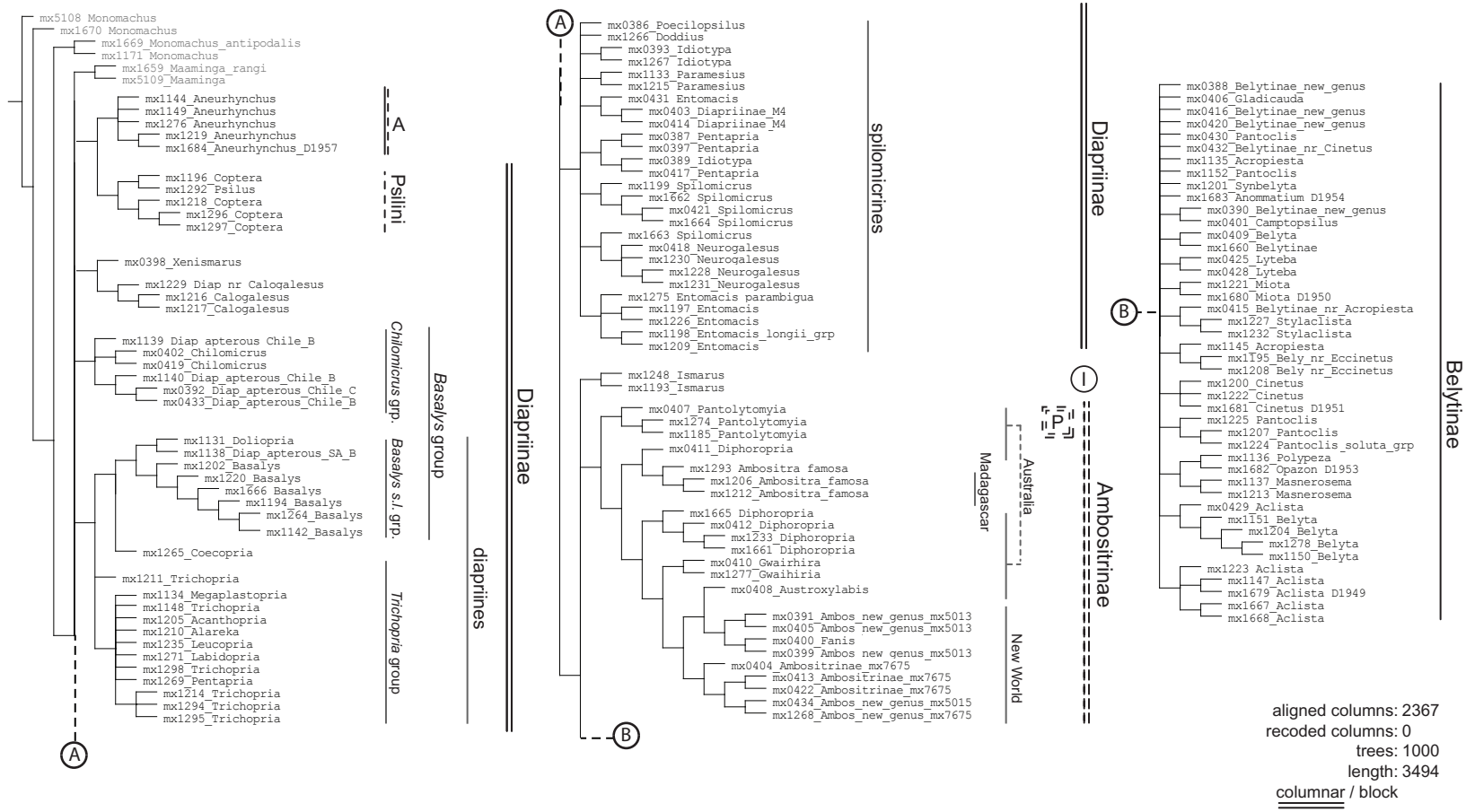

FIGURE 5.24. Results for partition 2 aligned with MAFFT.

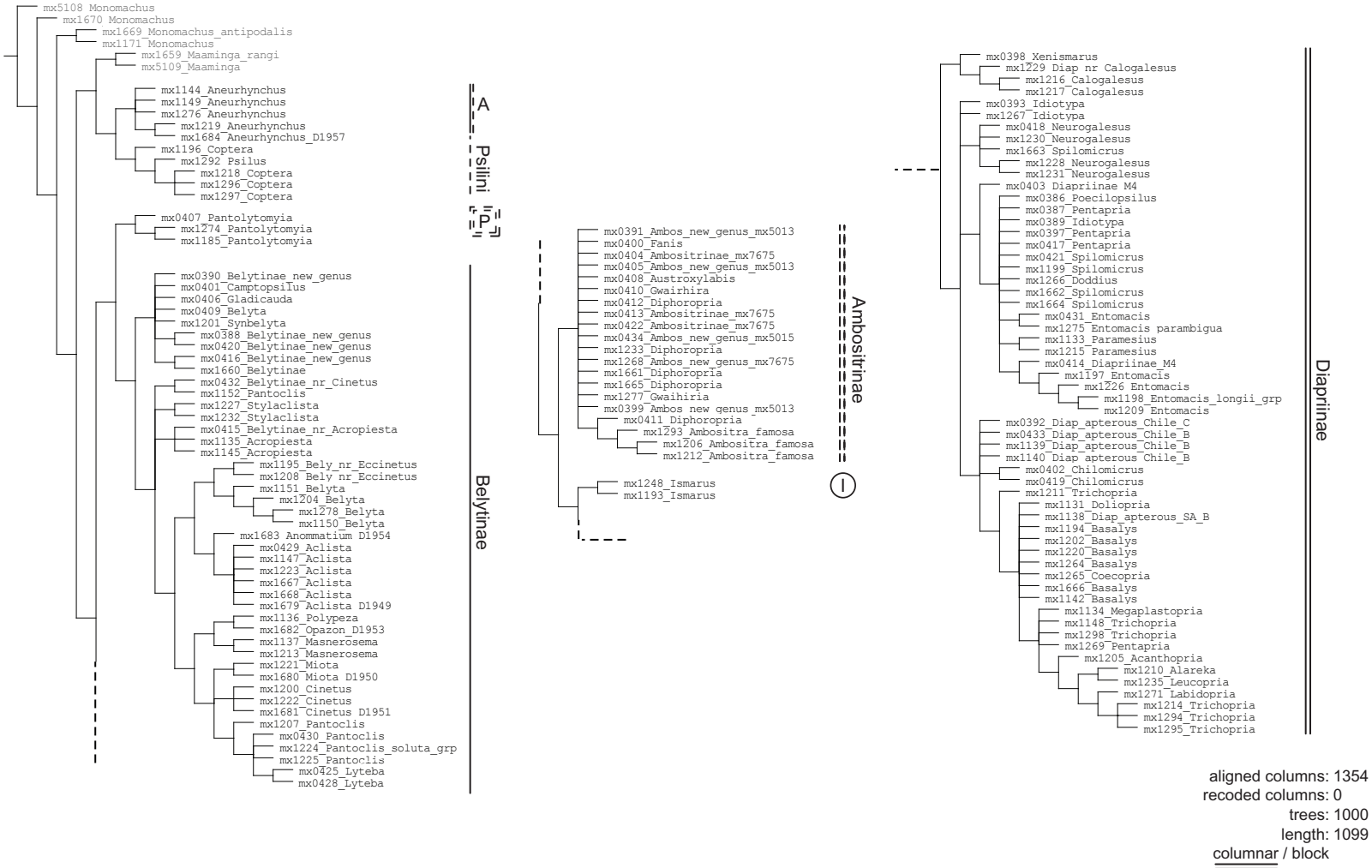

FIGURE 5.25. Results for partition 2 aligned with MAFFT with variable regions removed by gBlocks (Castresana, 2000). Note that with the exception of three clades (*Aneurhynchus*, Psilini, *Pantolytomyia*) the three major subfamilies are monophyletic.

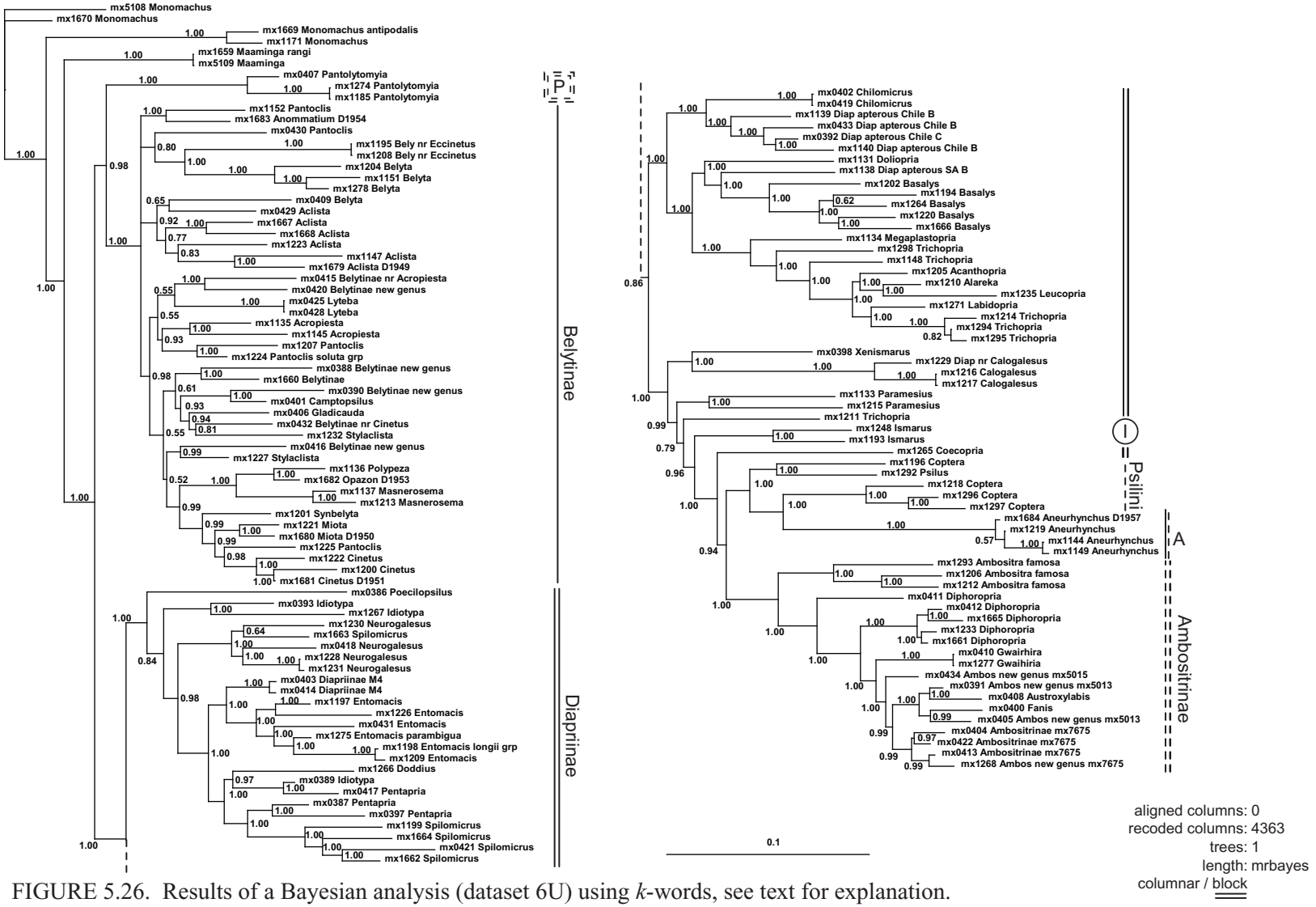

FIGURE 5.26. Results of a Bayesian analysis (dataset 6U) using *k*-words, see text for explanation.

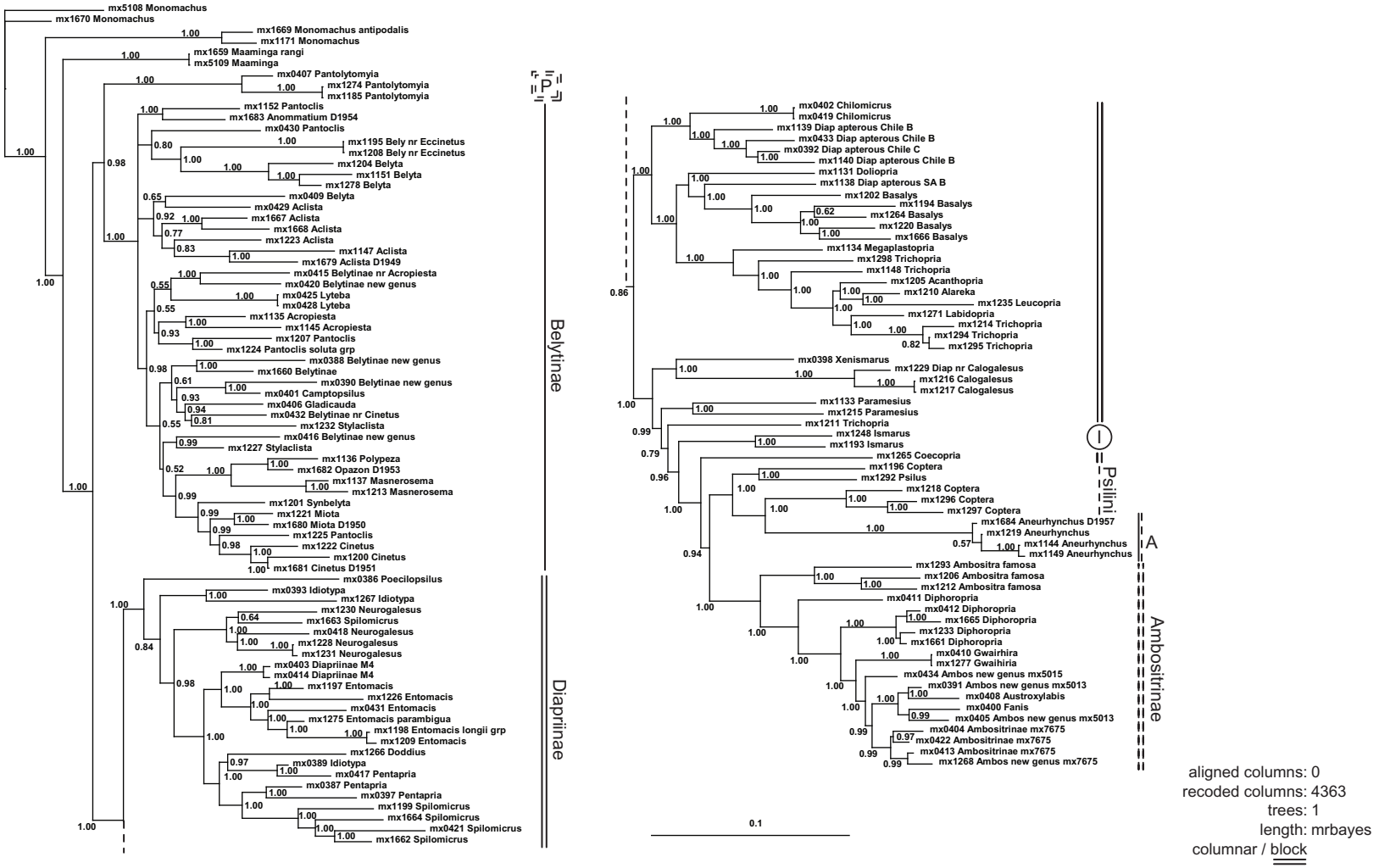

FIGURE 5.26. Results of a Bayesian analysis (dataset 6U) using *k*-words, see text for explanation.

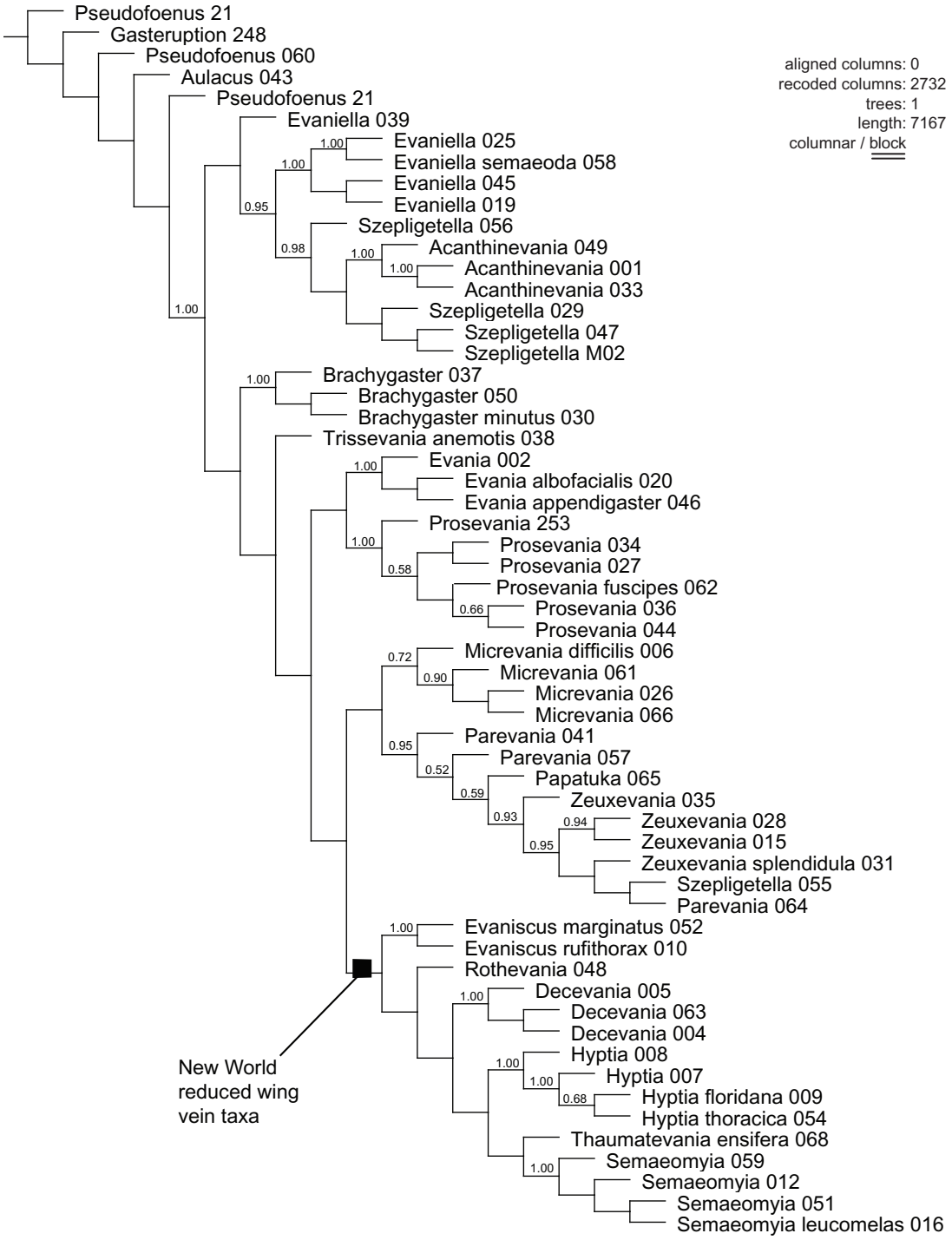

FIGURE 5.27. Results of a re-analysis of the 28S data alone from Deans et al. (2006). The data was unaligned and translated into *k*-words of size 8 before parsimony analyis. Posterior clade probabilities from that analysis (Fig. 6, Deans et al., 2006) are re-mapped onto this tree where clades were identical.

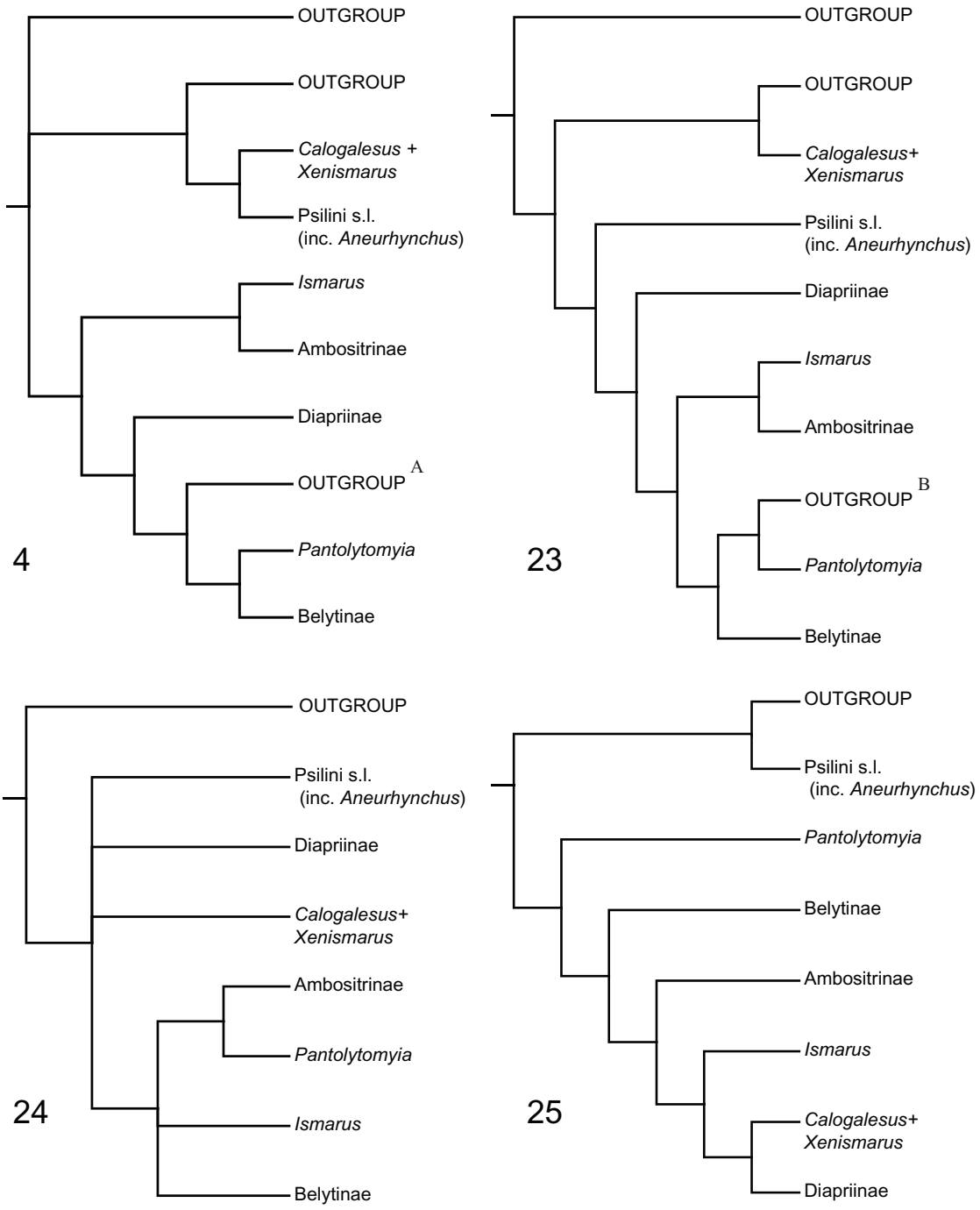

FIGURE 5.28. Summarized relationships from four analyses. Numbers reference previous figures from which summaries are dervied. [A]- Maamingidae and Monomachidae; [B]- Maamingidae.

The analyses presented here represent only a first-pass at understanding the consequences of hyper-partitioned approaches. While the results are generally promising they should not be considered as "proof" of the method, much further exploration is required. In particularly, stronger comparative criteria, perhaps coming in the form of additional phylogenies generated from novel data, are necessary. Various additional factors need to be accounted for in the future. For example, the effectiveness of each of the recoding methods is clearly tied to the degree of conservation of each of the three amplicons (18S < 28S D3-D5 < 28S D2). Our results do concretely illustrate that signal is present in ambiguously aligned regions (e.g. Fig. 5.9), and that this signal is congruent with data from other approaches.

Implications for Diapriid Systematics

The phylogenies generated have several important implications for diapriid systematics. Clades of note are indicated in Figure 5.4 with numbered circles, and throughout the figures vertical lines or symbols. These results are not necessarily in concordance with published concepts of intrafamilial groupings, based on morphology, yet many of them can be loosely reconciled with morphology in the process creating new hypotheses to test (see discussion below). We must emphasize that the following observations do not warrant nomenclatural decisions in our mind. Additional data, particularly morphological, will have to be reconciled with the hypotheses presented here before changes are invoked. Numerous untested hypotheses of relationship based on morphological characters exist (Naumann, 1982; Masner, 1993; Masner and García 2002), however to date there has been only a single quantitative phylogenetic analysis specifically addressing diapriid relationships (Loiácono and Margaria, 2000). Given this it is remarkable that with almost no exception all of the clades consistently recovered in our analyses have been previously hypothesized in the literature, typically in support of taxonomic decisions based on morphological evidence. There are, however, several novel relationships that were constantly recovered across a range of analyses.

We did not set out to specifically test the monophyly of Diapriidae nor seek to determine sister-group relationships for the family, however our results provide some information in this regard. Analysis of the 18S V4 amplicon alone (result not shown), a very conserved marker for the Hymenoptera, results in the monophyly of Monomachidae + Diapriidae + *Paleomymar* (Mymarommatidae). This relationships is also recovered in Dowton and Austin (2001) and Castro and Dowton (2006). Relationships inside this clade are largely collapsed though, with no support for monophyletic Diapriidae or Monomachidae. Note also that various critical outgroup taxa necessary for examining familial-level relationships are missing 18S in the present analysis (e.g. exemplars of Roproniidae, Pelecinidae, Heloridae), and furthermore there was no bootstrap support (1000 random additions sequences performed, data not shown) for this grouping. The sister to this clade is also ambiguous, being either Platygastroidea or Proctotrupidae.

In analyses that contained additional data, we for the most part assumed that Monomachidae+Maamingidae are sister to Diapriidae. In at least some of these analyses this resulted in polyphyly of the family (e.g. Fig. 5.9). Typically this was due to the position of the *Calogalesus*+*Xenismarus* clade, or Psilini+*Aneurhynchus* (= Psilini s.l., Notton, 2004). While both clades are notably morphologically aberrant for Diapriidae there is no strong morphological data suggesting that they should not fall within the family. The placement of Maamingids (Fig. 5.9) arising within Diapriidae could be due to the absence of 18S data for those taxa. A potentially more interesting result, based on evident morphological divergence, would have been the non-monophyly of Diapriidae with *Ismarus* being excluded, however we did not recover this result.

We were primarily interested in examining the hypotheses of subfamily monophyly and the relationships among those subfamilies. Only the Belytinae (and by definition the monobasic Ismarinae) were routinely recovered as monophyletic (sensu Notton, 2004), however this is based on the exclusion of *Aneurhynchus*, a genus that is variously placed in the Diapriinae and Beltyinae (Notton, 2004). The Ambositrinae, a group for which there is ample supporting morphological evidence were nearly always monophyletic with the exception of the placement of *Pantolytomyia*. Naumann (1982) recognized that species of *Pantolytomyia* were outliers relative to the remaining Ambositrinae. The sole analysis that recovered monophyly of the subfamily (Fig. 5.24)

was based on the algorithmic alignment, not the structural alignment. This may indicate problems with the structural alignment. The Diapriinae were typically recovered as monophyletic with the notable absence of two groups, the *Xenismarus+Calogalesus* clade, and the *Psilini*. The inability of analyses to recover a monophyletic Diapriinae is not completely unexpected. Masner (1993) notes that the family is traditionally defined on the basis of the absence or loss of characters. Based strictly on the analyses herein the potential crown lineages (see Fig. 5.28) are therefore: 1) *Xenismarus* plus *Calogalesus*; 2) Belytinae minus *Aneurhynchus*; 3) *Aneurhynchus* plus Psilini; 4) (remaining) Diapriinae; 5) Ambositrinae minus *Pantolytomyia*, 6) *Pantolytomyia* and 7) *Ismarus*. While it is difficult to prove (see Bergsten, 2005) that long branch attraction is occurring it is interesting to note that of the results listed above, groups 1), 3), and 6) all have proportionately long branches. Given the potential for these relationships to be biased by long branch attraction a conservative approach to interpretation of the results is taken here.

Relationships among the four subfamilies as traditionally defined remain confused given the present analyses. While cross analysis resolution is conflicting there appears to be at least some support for a sister relationships between the Ismarinae and Ambositrinae (Figs. 5.4-5.6, 5.12). The larger picture, however, is most often a grade, the polarity of which is invertible (Fig. 5.28).

The polyphyly of Ambositrinae was unexpected, and we expect that the inclusion of additional data will return a more consistently recovered monophyly. In several analyses where numerous MP trees were recovered a majority-rule consensus returned a monophyletic Ambositrinae with relatively high support (>70%). The relatively conservative strict-consensus approach combined with a small amount of signal for monophyly may be responsible for the general inability to recover a monophyletic Ambositrinae. Ambositrines are easily defined with an apparently strong apomorphy, the presence of a small second metasomal sternite (Masner 1961). Some corroborating morphological evidence for the independent placement of *Pantolytomyia* does exist, however, as species of *Pantolytomyia* have the ancestral number of free sternites and tergites, whereas remaining ambositrines exhibit various degrees of fusion (Naumann, 1982).

Naumann (1982) hypothesized the relationships among the known ambositrine genera spanning most of the southern hemisphere. Ambositrines are not known in the Palearctic and only a few species from two genera are known in North America. The molecular evidence largely supports his hypotheses (Naumann, 1982). With a few minor exceptions (notably the placement of *Austroxylabis*), three lineages were generally recognized ((Australian, Madagascar), New World) (membership identified in Fig. 5.24). Species of *Gwairhira* (Australia) were often intermediate between New and Old World groupings (e.g. Fig. 5.22). A recent effort to treat the New World genera of Ambositrinae has lead to the hypotheses of two additional major lineages (represented by terminals ending in "mx5013" and "mx7675"). These New World lineages are broadly defined by the presence or absence of the pronotal scrobe (see definition in Naumann, 1982). The existence of these two lineages is broadly supported by the molecular data (e.g. Figs. 5.5, 5.7-5.9, 5.12).

The Diapriinae is the most morphologically diverse subfamily of diapriids (Masner and García, 2002). Three tribes (Spilomicrini, Diapriini, Psilini) and three genera incertae sedis (*Coecopria, Peckidium, Calogalesus*) were identified by Masner and García (2002). We were able to sample all these groups except for *Peckidium*. Notton (2004) discusses the conclusions of Masner and García (2002), and concludes that two major clades, Psilini s.l. (including *Aneurhynchus*, Fig 5.4. "b'"), and Diapriini + Spilomicrini (Fig. 5.4, "d", "e" respectively for membership) can be defended. Our results concur with Notton's (2004) conclusions regarding Psilini, and with Masner and García (2002) uncertainty as to the placement of *Calogalesus*.

Notton (2004) tentatively includes the Psilini s.l. (including *Aneurhyncus* and *Labolips*) in Diapriinae, and provides a brief reexamination of the major pertinent characters. Notton (2004) concludes that Psilini are a good lineage of undetermined affiliation, with the possibility that it could be placed either in the Belytinae or Diapriinae. Masner and García (2002) excluded *Aneurhynchus* from the Diapriinae based on the presence of the belytine line, a longitudinal groove on the metasomal sternites. Notton (2004) suggests that this character is found in modified forms throughout the Diapriidae, and our independent observations concur with his hypothesis. Notton (2004) notes that venation characters unify *Aneurhynchus* plus Psilini, as does the presence of a

"macrotergite", the latter however is likely a plesiomorphy for the Diapriidae. Our present data most strongly suggest that Psilini s.l. are neither belytines nor diaprines, but represent their own lineage. The Psilini sensu Masner and García (2002) was included within the Diapriinae in only a single analysis (analysis 1O, result not shown). Most frequently Psilini occur as a sister to *Aneurhynchus* arising at or near the base of the Diapriidae (e.g. Figs. 5.4-6, 5.9, 5.11, 5.13, 5.25). Until stronger support for higher taxonomic relationships is recovered, however, it is premature to elevate the tribe to subfamily status.

The Spilomicrini, one of two major lineages with the Diapriinae, are largely recovered as monophyletic with several notable exceptions (see Fig. 5.23). The genus *Xenismarus* is rarely placed together with other members of the tribe (but see Figs. 5.9, 5.23), and most frequently found as sister to *Calogalesus*. This clade, *Xenismarus+Calogalesus* has not been proposed in the literature. The placement of this clade is extremely variable. Both genera do share apparently similar modifications of the posterior metasoma in females, with some telescoping possible. Analysis 14O (Fig. 5.13) does recover a *Xenismarus* and *Poecilopsilus* relationship, one suggested as possible in Masner and García (2002). Perhaps the most intriguing placement is illustrated in Fig. 5.25, a result using algorithmic alignment with no exclusion of data. In this case *Xenismarus+Calogalesus* are recovered as the basal members of a monophyletic Diapriinae minus Psilini. *Xenismarus* has a purportedly ancestral (Masner and García 2002) antennal formula (14-14), and this placement would be congruent with that hypothesis.

The other major exception is the placement of *Chilomicrus*. Described by Masner and García (2002) the genus also exhibits the 14-14 antennal formula, and is otherwise diagnosed vs. Spilomicrini sensu Masner and García (2002) by the presence of a basal vein oriented perpendicularly to the anterior margin of the forewing. *Chilomicrus* was consistently recovered as a basal member of Diapriini, most frequently in the *Basalys* group discussed further below.

Two major lineages were recovered for the Diapriini, the *Trichopria* and *Basalys* groups. This division has long been hypothesized based on morphological evidence (Masner and García, 2002), however the two lineages have not been formally treated as

separate entities. Masner and García (2002) believed the *Basalys* group would ultimately require tribal status. Our analyses suggest that *Chilomicrus* + Diapriini s. l. form a monophyletic clade (Figs. 5.4-5.6, 5.10, 5.11, 5.13, and see labels in Fig. 5.23) frequently sister to Spilomicrini. Elevation of the *Basalys* group to tribal status would render the Diapriini+*Chilomicrus* paraphyletic given some results (e.g. Fig 5.4, 5.5), though in others (e.g. Fig. 5.11) the group is monophyletic. At present we transfer *Chilomicrus* to a broadened concept of Diapriini (e.g. Fig. 5.4, 'e'), recognizing it as a sister taxon.

Recognition of the Diapriini+*Chilomicrus* relationship sheds light on several apparent morphological convergences between members of *Chilomicrus* and *Basalys*. The shared perpendicular basal vein, noted by Masner and García (2002) to be a highly derived apomorphy, can in fact be reinterpreted as plesiomorphic character for the expanded concept of Diapriini. Furthermore, numerous species of *Basalys* have a large number of morphological reductions (e.g. brachyptery, loss of notauli, minute mesoscutum). These species remain recognizable in females because of the presence of a massive 3-segmented clava. Species of *Chilomicrus* exhibit similar reductions, though with a greater range of variability. In the smallest members there is great convergence in form, such that generic placement becomes problematic. Results of the analyses here suggest that in Chile species of this form belong to *Chilomicrus*, despite some being tentatively identified (by MJY) as members of *Basalys* (Fig. 5.4, mx0433, mx1140). Minute members of *Basalys* are known world wide, and similar forms from South Africa (Fig. 5.4, mx1138) fall into the *Basalys* s.s. group.

An additional taxonomic clarification within the Diapriinae appears to be possible given our results. The genus *Coecopria*, incertae sedis in Masner and García (2002), was recovered nearly universally in the Diapriini. Masner (1969) suggested a potential relationship with species of *Doliopria*, and that conclusion is supported at least in part here (e.g. Fig. 5.5, 5.7, 5.9, 5.11, 5.12). *Coecopria* is variously sister to both the *Trichopria* and *Basalys* groups in our analyses. Regardless of its ultimate, placement it appears to fall well within the definition of the expanded concept of Diapriini, and as such can be removed from its present classification of Diapriinae incertae sedis.

Relationships within the definition of the Spilomicrini have only superficially been treated, primarily in the remarks of Masner and García (2002). Few consistent

relationships are evident in our analyses. Most notable is perhaps the placement of the genus *Paramesius*. *Paramesius* is variously placed as sister to remaining Spilomicrini (Fig. 5.23), or sister to the expanded concept of Diapriini (Fig. 5.4). In either case the genus appears to represent an intermediate case between Spilomicrini s.s. and Diapriini s.s. A proposed undescribed sister genus for *Entomacis*, "M4" was included in our analyses. This relationship was recovered in several analyses (e.g. Figs. 5.4, 5.5). Relationships of *Entomacis* to *Poecilopsilus* and *Doddius* (Fig. 5.10), proposed by Masner and García (2002) were also recovered, though not in all cases (Fig. 5.4).

Relationships in the remaining subfamily, Belytinae, are difficult to interpret. The subfamily is the most poorly understood taxonomically, and the present results based on molecular evidence appear to do little to clarify this problem. Few relationships recovered for small groups of genera (e.g. two or three) are in congruence with interpretations based on morphology (e.g. Fig. 5.23 *Cinetus* + *Miota*), however, many relationships group markedly different morphologies (e.g. *Aclista* + *Belyta*), and a majority of clades can not be matched to presently held higher classifications (while there is little proposed in this regard see Macek, 1989ab, 1995) .

Given the present results it should be noted that with few exceptions a majority of the relationships recovered in these analyses were previously hypothesized based on morphological work. The addition of molecular data has helped to clarify some relationships and raised several interesting questions, particularly those pertaining intrageneric relationships within the Belytinae. While taxon sampling in the present analysis included members of all the major lineages, several key taxa remain to be sampled. In particular additional members of the Psilini s.l. (e.g. *Ortona*, *Aneuropria*, *Labolips*), the engimatic *Peckidium* Masner and García, and several known lineages of Australian and African myrmecophiles are desirable additions. It is important to note that the scope of the analyses undertaken here is a small fraction of what can be attempted. In this regard Bayesian, likelihood, and direct-optimization approaches may provide further insight those relationships that remained clouded. Finally, with respect to molecular markers it is clear that the 28S D2 expansion segment will be highly informative for diapriids, it appears to contain much more signal than either of the other two amplicons used here.

Discussion

Significant debate rages about the utility of "structural" alignments. For example Wheeler et al. (2006) finish their discussion on alignment with "In light of the many epistemological and theoretical shortcomings of the approach, manual alignment should be seen as a nonscientific procedure that is best avoided." This is clearly a gross oversimplification. Similar broad claims by proponents of structural alignment (structural alignments are manual), for instance that structural alignments result in more accurate hypotheses of phylogeny (Kjer, 2004), are equally untested (but see summary in Gillespie, 2004) oversimplifications. To date there has been little empirical testing of either of these claims, and indeed it is unclear as to the arena wherein both could compete. At least part of the problem occurs in interpreting a structural alignment (or hyper-partitioned alignment) as the fixed end product from which phylogenies are derived, this is once again a simplification that should be avoided. Structural alignments should be seen simply as maximally useful starting points. For example, POY analyses can be derived from a structural alignment, but not vice versa. Due to the additional information in structural alignments (e.g. isolation of stems and loops) many additional questions can be asked, starting with a broad approach "what are the consequences of secondary structure to phylogenetic analysis?" To answer this question explicit ideas about the structural properties of a molecule must be encoded in our starting points. Algorithmic approaches that ignore these properties greatly weaken the potential conclusions they can make about "the consequences of secondary structure". More specific questions, such as "what are the consequences of basepairing?" can also be envisioned. Most algorithmic approaches presently employed to provide alignments for phylogenetic inference, including the dynamic approach used in POY, fail to address this question. Wheeler et al. (2007) recognize that a character is "a historically independent transformation series". This assumption is simply not the case (Table 5.1) for many genes important to phylogenetic inference. Structural approaches allow us to identify such problems, and as such are an important first step prior to analyzing data.

Wheeler and Honeycutt (1988) suggested that loop regions held more information than pairing regions, and that pairing regions might positively mislead (contain

"disinformation content"). This is another example of a question that can only be addressed using structure to guide the delimitation of structural elements. Our results suggest that loop regions do indeed contain phylogenetic signal, however it less clear whether pairing regions positively mislead.

From a pragmatic standpoint, if the overall goal is simply a reasonably derived phylogeny, then the structural approach as presented here can be sharply criticized in several regards. Figures 5.24 and 5.25 represent "traditional" algorithmic approaches, efforts that took orders of magnitude less time to generate (i.e. structural multiple-sequence alignment took several weeks to complete). In this study the trees illustrated in Figs. 5.24 and 5.25 are as congruent as the structural approaches, if not more so, with proposed relationships based on morphological data. Given morphological congruence as a criterion with which we compared the various methods, the conclusion of this empirical study should be that algorithmic approaches are to be used, with the saved effort reapplied to gathering additional data, or incorporating morphological data. This is perhaps a harsh conclusion, however it is one that can help work towards generating other, perhaps more meaningful criteria with which to test results with. A second major critique comes following the analysis when new data are added. At present there are few efficient mechanisms that can reflect what is learned during the course of analysis (e.g. the problem illustrated in Fig. 5.22) *back into* the alignment. Software tools, and a more rigorous protocol for finding and implementing this information are needed. The alignment presented here is small to medium sized in today's standards, nevertheless the effort required to annotate it was very large. As noted in Ogden et al. (2005) these types of efforts will not scale with the exponentially increasing amounts of data. Finally, it is well known that differences in parameter settings result in differing topologies, particularly in length heterogeneous data (e.g Morrison and Ellis, 1997). Morrison and Ellis (1997) conclude that primary differences in topology are realized by differences in alignment, and perhaps not by the method used. Though manual alignments are not parameterized per se, and a not unexpected obvious result (thow perhaps not often empirically illustrated) we have shown that they too can result in differing topologies, even under a single criterion (i.e. parsimony). While this is not a completely unexpected results (the homology statements are different), it is not often empirically illustrated.

While there is noticeable congruence among our results, there is also a great degree of variation.

If the criterion for "success" is more than a congruent phylogeny then many additional outcomes of an empirical study such as this can be presented as positive results. The idea of hyper-partitioned alignments forces the question "what is a partition?". Because of the small size of partitions in hyper-partitioned alignments metrics like the ILD are likely to fail, and it is unclear as to how Bayesian or likelihood models could be chosen and applied to these data as well. This raises the question as to whether there are meaningful limits to partition size. The strings of data which we analyze have biological functions, and ultimately we would like to be able to say something about those functions. Algorithmic approaches alone will rarely lead to meaningful conclusions about these functions. Finally, thinking about ways in which to capture data in ambiguously aligned regions has lead indirectly to the *k*-word approaches attempted here. Empirically this approach has thus far done remarkably well, but much further study is needed.

While there is much research in the development of tools that allow for the alignment and generation of rRNA models the number of tools that are specifically oriented towards phylogenetic analyses of large scale datasets remains minimal. We can conclude then that new annotation and automation tools are a high priority, particularly those that can handle and report complex structural meta-data similar to that discussed here. Propellar Twist (http://bioinformatics.org/paradise/) and jPHYDIT (Jeon et al., 2005) are promising ongoing efforts in this regard. Following hyper-partitioning analyses in a likelihood framework are also desirable, integrating alignment data such as presented here into platforms like HyPhy (http://www.hyphy.org/) may facilitate this process.

Static (as opposed to dynamic?!) alignments are undesirable from both from a construction standpoint and from an analytical one. Ideally one would like to be able to add data quickly as it becomes available, and analyze it continuously. Molecular evolution does not occur in the fixed state that a multiple-sequence alignment implies. Building phylogenies based on explicit models of rRNA evolution that allow for insertion and deletion events (e.g. Holmes, 2004, 2005) are also desirable, though these approaches

are in their relative infancy. It is hoped that both the models presented here, and the tools with which to handle the data, will act as a basis for larger, more dynamic analyses and more rigorous testing of the claims made by those who produce hyper-partitioned or structural analyses.

Acknowledgements

This project was funded by NSF PEET grant DEB-0328922 to Robert Wharton. Joe Gillespie helped in the preparation of this paper. Some critical sequence data were generously provided by Debra Murray (data from HymATOL NSF grants EF-0337220) and by the Heraty Lab (University of California, Riverside). Robert Wharton, Jim Woolley, and Rodney Honeycutt (TAMU) and Anthony Cognato (TAMU/MSU) provided innumerable helpful suggestions. Thanks to Jacques Dubois (TAMU, USNM), for various insightful discussions on the character coding methods and issues pertaining to the ILD metric. Andy Deans provided helpful comments on the *k*-word approach, and aided in the reanalysis of his data. Dave Adelson (TAMU) provided computing support at various stages of this project. Critical specimens were obtained through granted efforts to Brian Fisher and Margaret and Al Newton. Thanks to Lesley Washburn for help with sequencing, Lubomir Masner for discussion of all things diapriid, and James Munro for technical help with TNT.

CHAPTER VI
CONCLUSION

Biological systematists face a world where information accumulates logarithmically. Not only are these scientists required to participate in the generation of this information, but they are expected also to master the growing number of methods available to processes and disseminate it. This situation often leads young systematists to become jacks-of-all-trades but masters-of-none, which has both positive and negative consequences.

Advances in technologies that generate and disseminate descriptive data (e.g. DNA sequencing, the World Wide Web) equate to new tools that must be mastered by taxonomists whose traditional job has been largely the study of the whole organism from the morphological and classificatory perspectives. In the recent past a professional taxonomist could spend a lifetime intimately connected solely with his or her organism(s) of interest. Today at least some of that intimacy is lost due to the increasing demands placed on the taxonomist to learn new technologies (e.g. imaging, sequencing, databasing). One can mitigate some of these demands by collaboration, and mx, and the molecular work presented herein, is an example of such an approach. The necessity of collaboration represents a paradigm shift for systematists, who frequently were the lone experts for their taxon of study.

Lone taxonomists, excommunicated to the darkened corners of museums, presumably will be relegated to hoary curiosities with the rise of collaborative mechanisms enabled (primarily) by the World Wide Web. The melding of digital mechanisms and taxonomy has lead to the advent of the "cybertaxonomist" (e.g. see Vince Smith’s blog; a Google™ search will find it long after any provided URL would become broken). While Smith states that "we are all cybertaxonomists now" I argue that not until we are more *efficient* at doing what we're doing (diagnosing, describing, keying, imaging, etc.) are we truly cybertaxonomists, or at least at that point cybertaxonomy will be a meaningful proposition.

Leveraging efficiencies from digital technologies is not trivial. Taxonomists must still progress through all the basic stages necessary for proper comparative and descriptive work, i.e. no efficiency has yet been recognized that will eliminate the depth of information a trained taxonomist can deliver after long periods of careful study. Though at its core taxonomy seems to be a daunting task, most taxonomists tackle the process with a fervent ardour.

Zealously undertaking a job as a taxonomist (or systematist) is done for good reasons. There is seemingly endless biodiversity remaining to be described (in particular within the parasitic Hymenoptera). The high probability of discovering species new to science continues to drive taxonomists, who often spend their lifetimes deeply focused on their trade, ever anticipating the next thrill. With respect to diapriids it is clear that the addition of new taxa to the phylogenetic framework presented here will result in novel and important consequences for both diapriid taxonomy and Hymenoptera phylogenetics as a whole. At some point, however, adding new taxa to a phylogeny, or analyzing old data with a new method, will not be the most efficient way to "raise the bar".

Finding a balance between the constancy demanded in the comparative approach to describing new species and taxa, and the learning and applying of new methods to new (or old) data is *de rigueur* for any systematist. Given that there seems to be no sign of slow-down in the development of methods (i.e. those with which to generate new information) for taxonomists, striking a balance between training one's self and collecting/analyzing data may be difficult if not already impossible. Attempts to master all trades (e.g. alpha taxonomy, quantitative phylogenetics, informatics, monography, languages) will undoubtedly fail, and becoming a master-of-one will likely lead one to overlook the efficiencies necessary to address the taxonomic impediment.

This dissertation is largely an attempt at reconciling "traditional" taxonomic practices with the burgeoning new methods largely encompassed by the umbrella of "Biodiversity Informatics". Paradoxically (though not unexpected, as noted above), the time required for this reconciliation has directed effort away from actually revising new taxa. For instance many Ph.D. students typically treat (i.e. describe or re-describe) upwards of a 100 if not more taxa during the course of their studies. This seeming disparity then, for a dissertation that initially was planned as a relatively pure taxonomic

endeavor, can perhaps be justified as follows.

When new technologies are envisioned or described there is typically a significant lag between their inception and their usefulness in any widespread capacity. I hope that the underlying infrastructure developed here will lead to a spike in the efficiency required to treat the remaining Diapriidae in reasonable timeframe (i.e. under 100 years). This is by no means possible by one person, but rather by a team of collaborators, bound by mechanisms made possible by the World Wide Web. Databases, such as mx, which aspire to at least in part represent these mechanisms, are for all intents and purposes in their infancy. They are born out of necessity in response to increasing demands on the taxonomist, but are by no means yet at a stage where they become the taxonomists third hand. Only when databases and molecular phylogenies are not but an afterthought to the taxonomist/systematist will we be at the stage where the true efficiencies needed to treat the taxonomic impediment are met.

REFERENCES

Adoutte, A., G. Balavoine, N. Lartillot, O. Lespinet, B. Prud'homme, and R. de Rosa. 2000. The new animal phylogeny: Reliability and implications. Proc. Natl. Acad. Sci. 97:4453-6.

Aguiar-Menezes, E. L., E. B. Menezes, P. S. Silva, A. C. Bittar, and P. C. R. Cassino. 2001. Native hymenopteran parasitoids associated with *Anastrepha* spp (Diptera:Tephritidae) in Seropedica City, Rio de Janeiro, Brazil. Fla. Entomol. 84:706-711.

Aguiar-Menezes, E. L., E. B. Menezes, and M. S. Loiácono. 2003. First record of *Coptera haywardi* Loiácono (Hymenoptera: Diapriidae) as a parasitoid of fruit-infesting Tephritidae (Diptera) in Brazil. Neotropica 32:355-358.

Aliniazee, M. T. 1975. Susceptibility of diapausing pupae of Western Cherry Fruit fly (Diptera:Tephritidae) and a parasite (Hymenoptera:Diapriidae) to subfreezing temperatures. Environ. Entomol. 4:1011-1013.

Angelov, S., B. Harb, S. Kannan, S. Khanna, and J. Kim. 2006. Efficient enumeration of phylogenetically informative substrings. Lecture Notes in Computer Science vol. 3909.

Ashmead, W. H. 1893. A monograph of the North American Proctotrypidae. Bulletin of the United States National Museum, 45:1-472.

Ashmead, W. H. 1902. A new genus of diapriids from Texas. Biol. Bull., Marine Biological Laboratory, Woods Hole, MA 3:15.

Backlund, H. 1945. Wreck fauna of Sweden and Finland, ecology and chorology. Opuscula Entomologica Supplementum 5:1–237.

Baeza-Larios, G., J. Sivinski, T. Holler, and M. Aluja. 2002. The ability of *Coptera haywardi* (Ogloblin) (Hymenoptera: Diapriidae) to locate and attack the pupae of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann) (Diptera: Tephritidae), under seminatural conditions. Biol. Control 23:213-218.

Bailey, N. S. 1947. *Trichopria tabanivora* Fouts in Massachusetts (Hymenoptera: Diapriidae). Psyche 54:142-144.

Baker, R. H., and R. DeSalle. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46:654–673.

Baker, A. C., W. E. Stone, C. C. Plummer, and M. McPhail. 1944. A review of studies on the Mexican Fruitfly and related Mexican species. United States Department of Agriculture Miscellaneous Publication 531:1-155.

Baker, G. L. 1978. The biology of a species of *Doleschalla* (Diptera: Tachinidae), a parasite of *Pantorhytes szentivanyi* (Coleoptera: Curculionidae). Pac. Insects 19:53-64.

Ban, N., P. Nissen, J. Hansen, P.B. Moore, and T. A. Steitz. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 289:905-920.

Baranowski, R., H. Glenn, and J. Sivinski. 1993. Biological control of the Caribbean Fruit Fly (Diptera:Tephritidae). Fla. Entomol. 76:245-251.

Baudoin, R. 1949. Un Hymenoptère nouveau pour la zone intercotidale de Roscoff. In XIIIe Congrès International de Zoologie, Paris. pp 69-470.

Baudoin, R. 1952. Contribution à l'oecologie et à la biogéographie de la zone intercotidale des Charentes. C.R. sommaires Biogéographie 251:73-82.

Belshaw R., and D. L. J. Quicke. 2002. Robustness of ancestral state estimates: Evolution of life history strategy in ichneumonoid parasitoids. Syst. Biol. 51:450-477.

Berendson W. G. 1995. The concept of potential taxa in databases. Taxon 44:207-212.

Berendsohn, W. G., A. Güntsch, and D. Röpert. 2003. Survey of existing publicly distributed collection management and data capture software solutions used by the world's natural history collections. The Global Biodiversity Information Facility, Copenhagen. http://circa.gbif.net/Public/irc/gbif/digit/library.

Bergsten, J. 2005. A review of long-branch attraction. Cladistics 21:163-193.

Blume, R. R. 1984. Parasites of Diptera associated with bovine droppings on a pasture in east central Texas. Southwest. Entomol. 11:215-222.

Bradley, S. W., D. C. Booth, and D. C. Sheppard. 1984. Parasitism of the Black Soldier Fly by *Trichopria* sp (Hymenoptera: Diapriidae) in poultry houses. Environ. Entomol. 13:451-454.

Brown, J. W. 1999. The Ribonuclease P Database. Nucl. Acids Res. 27:314.

Brown, H. P., and R. Arrington. 1967. A new species of *Psephenus* from Mexico and Texas (Coleoptera, Dryopoidea, Psephenidae). Coleopterists' Bulletin, 21:86-91.

Brown, H. P. 1967. Psephenids parasitized by eulophid and diapriid wasps. Amer. Zool. 7:193.

Brown, H. P. 1968. *Psephenus* (Coleoptera: Psephenidae) parasitized by a new Chalcidoid (Hymenoptera: Eulophidae) II. Biology of the Parasite. Ann. Entomol. Soc. Am. 61:452-456.

Brown, H. P. 1987. Biology of riffle beetles. Ann. Rev. Entomol. 32:253-273.

Brues, C. T. 1908. Three hymenopterous parasites of the cabbage maggot. State Entomologist of Minnesota 12:192-196.

Brues, C. T. 1916. A new species of *Lepidopria* from North America. Psyche 23:126-127.

Buckingham, G. R. 1989. *Lemnaphila scotlandae* (Diptera: Ephydridae) and three of its parasites discovered in Florida. Florida Entomol. 72:219-220.

Cameron, A. E. 1926. Bionomics of the Tabanidae (Diptera) of the Canadian prairie. Bull. Entomol. Res. 17:1-42.

Cameron, J. P., and A. W. Morrison. 1974. *Psilus* sp (Hyemnoptera: Diapriidae), a parasite of the pupal stage of the Apple Maggot *Rhagoletis pomonella* (Diptera: Tephritidae) in southwestern Quebec. Phytoprotection 55:13-16.

Campbell, B., J. Heraty, Y. Rasplus, K. Chan, J. Steffen-Campbell, and C. Babcock. 2000. Molecular systematics of the Chalcidoidea using 28S-D2 rDNA. pp.59-73 In AD Austin, M Dowton (Eds), Hymenoptera: Evolution, biodiversity and biological control, CSIRO, Collingwood, Victoria, 512 pp.

Cannone, J. J., S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Müller, N. Pande, Z. Shang, N. Yu, and R.R. Gutell. 2002. The comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BioMed Central Bioinformatics. 3:15.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540-52.

Castro, L., and M. Dowton. 2006. Molecular analyses of the Apocrita (Insecta: Hymenoptera) suggest that the Chalcidoidea are sister to the diaprioid complex. Invert. Syst. 20:603–614.

Cate, J. H., M. M. Yusupov, G. Z. Yusupova, T. N. Earnest, and H. F. Noller. 1999. X-ray crystal structures of 70S ribosome functional complexes. Science 285:2095–2104.

Chambers, V. H. 1955. Some hosts for *Anteon* spp (Hymenoptera, Dryinidae) and a hyperparasite *Ismarus* (Hymenoptera, Belytinae). Entomol. Mon. Mag. 91:114-115.

Chambers, V. H. 1971. Large populations of Belytinae (Hymenoptera, Diapriidae). Entomologist's Monthly Magazine, 106:149-154.

Chambers, V. H. 1981. A host for *Ismarus halidayi* Foerst (Hym, Diapriidae). Entomol. Mon. Mag. 117:29.

Chapus, C., C. Dufraigne, S. Edwards, A. Giron, B. Fertil, and P. Deschavanne. 2005. Exploration of phylogenetic data using a global sequence analysis method. BMC Evol. Biol. 5:63.

Clausen, C. P., D. W. Clancy, and Q.C. Chock. 1965. Biological control of the Oriental Fruit Fly (*Dacus dorsalis* Hendel) and other fruit flies in Hawaii. Tech. Bull. U.S.D.A., Bur. Ent. Plant Quarantine, pp 1-102.

Combs, R. L., and C.E. Hoelscher. 1969. Hymenopterous pupal parasitoids found associated with the Horn Fly in Northeast Mississippi. J. Econ. Entomol. 62:1234-1235.

Coon, B. R. 2000. Biology of *Trichopria columbiana* (Hymenoptera: Diapriidae), an endoparasitoid of *Hydrellia pakistanae* (Diptera: Ephydridae) a biological control agent of the aquatic weed *Hydrilla verticillata* (Hydrocharitaceae). Entomology and Nematology Department, University of Florida, Gainesville, Florida.

Cramér, H. 1999. Mathematical methods of statistics, Princeton University Press, Princeton, New Jersey. pp 1-575.

Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in Phylogenetic Analysis. Mol. Biol. Evol. 12:814-822.

Dallwitz M. J., 1980. A general system for coding taxonomic descriptions. Taxon 29:41–46.

Dallwitz, M. J., and T. A. Paine. 2004. Definition of the Delta format. http://deltaintkey.com/www/standard.htm.

Dallwitz, M. J., T. A. Pain, and E. Zurcher. (2005) Principles of interactive keys. http://delta-intkey.com/www/ interactivekeys.htm.

Deans, A. R. 2005. Annotated catalog of the world's ensign wasp species (Hymenoptera: Evaniidae). Contributions of the American Entomological Institute 34:1-164.

Deans, A. R., and M. J Yoder. 2006. Hymenoptera Glossary. http://hymglossary.tamu.edu.

Deans, A. R., J. J. Gillespie, and M. J. Yoder. 2006. A phylogenetic evaluation of ensign wasp classification (Hymenoptera: Evaniidae) based on molecular data and insights from ribosomal secondary structure. Syst. Ent. 31:517-528.

De Rijk, P., and R. De Wachter. 1993. DCSE, an interactive tool for sequence alignment and secondary structure research. Bioinformatics 9:735-740.

Deschavanne, P. J., A. Giron, J. Vilain, G. Fagot, and B. Fertil. 1999. Genomic signature: Characterization and classification of species assessed by chaos game representation of sequences. Mol. Biol. Evol. 16:1391-1399.

Dettai, A., N. Bailly, R. Vignes-Lebbe, and G. Lecointre. 2004. Metacanthomorpha: Essay on a phylogeny-oriented database for morphology--the acanthomorph (teleostei) example. Systematic Biology 53:822-34.

Dowell, R. D., and S. R. Eddy. 2006. Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints. BMC Bioinformatics 7:400.

Dowton, M., A. D. Austin, N. Dillon, and E. Bartowsky. 1997. Molecular Phylogeny of the apocritan wasps: The Proctotrupomorpha and Evaniomorpha. Systematic Entomology 22:245-255.

Dowton, M., and A. D. Austin. 2001. Simultaneous analysis of 16S, 28S, COI and morphology in the Hymenoptera: Apocrita - evolutionary transitions among parasitic wasps. Biological Journal of the Linnean Society 74:87-111.

Dodd, A. P. 1920. Two new Hymenoptera of the superfamily Proctotrypidae from Australia. Proc. Linn. Soc. N.S.W. 45:443-446.

Dodd, A. P. 1930. New Hymenoptera Proctotrypoidea from Victoria. Proc. R. Soc. Victoria 43:26-35.

Early, J. W. 1978. New Diapriinae (Hymenoptera: Diapriidae) from the South Island and subantarctic islands of New Zealand. J. Roy. Soc. N.Z. 8:207-228.

Early, J. W. 1980. The Diapriidae (Hymenoptera) of the southern islands of New Zealand. J. Roy. Soc. N.Z. 10:153-171.

Early, J. W., and D. S. Horning. 1978. Two new wasp parasites (Hymenoptera: Diapriidae) of New Zealand Sciomyzidae (Diptera). J. Roy. Soc. N.Z. 8:231-237.

Early, J. W., and I. D. Naumann. 1990. *Rostropria*, a new genus of opisthognathous diapriine wasp from Australia, and notes on the genus *Neurogalesus* (Hymenoptera: Proctotrupoidea: Diapriidae). Invert. Taxon. 3:523-550.

Early, J. W., L. Masner, A. D. Austin, and I.D. Naumann. 2001. Maamingidae, a new family of proctotrupoid wasp (Insecta: Hymenoptera) from New Zealand. Invertebrate Taxonomy 15:341-352.

Edgar, R. C. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

Edgar, R. C., and S. Batzoglou 2006. Multiple sequence alignment. Current Opinion in Structural Biology 16:368-73.

Erwin, T. L. 2000. Naming species, a new paradigm for crisis management in taxonomy: Rapid journal validation of scientific names enhanced with more complete descriptions on the internet. Coleopterists' Bulletin 54:269-278.

Fahringer, J. 1922. Contributions to a knowledge of the habits of some parasitic hymenoptera with special regard to their importance in the biological control of injurious insects. Zeitschrift fuer Angewandte Entomologie 8:325-388.

Fernández-Marín, H., J. K. Zimmerman, and W.T. Wcislo. 2006. *Acanthopria* and *Mimopriella* parasitoid wasps (Diapriidae) attack *Cyphomyrmex* fungus-growing ants (Formicidae, Attini). Naturwissenschaften, Berlin 93:17-21.

Ferriere, C. 1927. Parasites de *Perrisia pyri*, Bouche (Dipt, Cecidom). Bull. Entomol. Res. 17:421-422.

Figg, D. E., R. D. Hall, and G. D. Thomas. 1983. Insect parasites associated with Diptera developing in bovine dung pats [of] Central Missouri pastures. Environ. Entomol. 12:961-966.

Floate, K. D., B. G. Khan, and G. Gibson. 1999. Hymenopterous parasitoids of filth fly (Diptera:Muscidae) pupae in cattle feedlots. Can. Ent. 131:347-362.

Floate, K. D., T. J. Lysyk, G. A. P. Gibson, and T. D. Galloway. 2002. Ch *37 Musca domestica* L, House fly (Diptera: Muscidae). Pages 190-195 *in* Biological control programmes in Canada 1981-2000 (P. A. Mason, G. A. P. Gibson, K. D. Floate, T. J. Lysk, G. A. P. Gibson and T. D. Galloway eds.). CABI Publishing, Wallingford, UK.

Foerster, A. 1856. Hymenopterologische Studien. II. Heft. Chalcidae und Prototrupii. Ernst ter Meer, Aachen, 152 pp.

Fouts, R. M. 1939. Descriptions of one new genus and three new species of Diapriidae (Hymenoptera). Proc. Entomol. Soc. Wash. 41:260-263.

Froggatt, W. W. 1917. Sheep maggot flies. Farmers' Bulletin 113:32-33.

Froggatt, W. W. 1918. Sheep maggot flies. Farmers' Bulletin 122:18.

Fullaway, D. T. 1918. Division of Entomology. Hawaiian Forester and Agriculturist 15:381-382.

Garbarczyk, H. 1981. Tybelaki (Proctotrupoidea, Hymenoptera). Fragmenta Faunistica, 26:325-340.

García, F. R. M., and E. Corseuil. 2004. Native hymenopteran parasitoids associated with fruit flies (Diptera:Tephritidae) in Santa Catarina State, Brazil. Fla. Entomol. 87:517-521.

García, J.-L., and R. Montilla. 2001. *Coptera haywardi* Loiácono (Hymenoptera: Diapriidae) parasitoide de pupas de *Anastrepha* spp (Diptera: Tephritidae) en Venezuela. Entomotropica 16:191-195.

Gardner, P. P., and R. Giegerich. 2004. A comprehensive comparison of comparative RNA structure prediction approaches. BMC Bioinformatics. 5:140.

Gardner, P. P., A. Wilm, and S. Washietl. 2005. A benchmark of multiple sequence alignment programs upon structural RNAs. Nucleic Acids Res. 33:2433–2439.

Ghorpade, S. A., S. S. Jadhav, and D. S. Ajri. 1982. Record of *Trichopria* sp (Hymenoptera: Diapriidae)--a parasitoid of rhizome fly, macquart in Maharashtra *Mimegralla coeruleifrons*, pest of turmeric and ginger, India. Journal of Maharashtra Agricultural Universities 7:193.

Gibson, G. A. P. 1985. Some pro- and mesothoracic characters important for phylogenetic analysis of Hymenoptera, with review of terms used for structures. Canadian Entomologist, 117:1395-1443.

Gibson, G. A. P. 1999. Sister-group relationships of the Platygastroidea and Chalcidoidea (Hymenoptera) - an alternate hypothesis to Rasnitsyn (1988). Zoologica Scripta, 28:125-138.

Gillespie, J. J. 2004. Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. Mol. Phylogenet. Evol. 33:936-943.

Gillespie, J. J., J. J. Cannone, R. R. Gutell, and A.I. Cognato. 2004. A secondary structural model of the 28S rRNA expansion segments D2 and D3 from rootworms and related leaf beetles (Coleoptera: Chrysomelidae; Galerucinae). Insect Mol. Biol. 13:495-518.

Gillespie, J. J., C. H. McKenna, M. J. Yoder, R. R. Gutell, J. S. Johnston, J. Kathirithamby, and A.I. Cognato. 2005a. Assessing the odd secondary structural properties of nuclear small subunit ribosomal RNA sequences (18S) of the twisted-wing parasites (Insecta: Strepsiptera). Insect Molecular Biology 14:625-643.

Gillespie, J. J., M. J. Yoder, and R. A. Wharton. 2005b. Predicted secondary structures for expansion segments D2-D10 of the 28S large subunit ribosomal RNA from ichneumonoid Hymenoptera: Homology assignment and phylogenetic implications. J. of Molecular Evolution 61:114-137.

Gillespie, J. J., J. B. Munro, J. M. Heraty, M. J. Yoder, A. K. Owen, and A. E. Carmichael. 2005c. A secondary structural model of the 28S rRNA expansion segments D2 and D3 for chalcidoid wasps (Hymenoptera: Chalcidoidea). Molecular Biology and Evolution 22:1593-1608.

Gillespie, J. J., J. S. Johnston, J. J. Cannone, and R. R. Gutell. 2006. Characteristics of the nuclear (18S, 5.8S, 28S, and 5S) and mitochondrial (16S and 12S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): Structure, organization and retrotransposable elements. Insect Mol. Biol. 15:657-686.

Gourlay, E. S. 1930. Preliminary host-list of the entomophagous insects of New Zealand. New Zealand Department of Science and Industry Research, Bulletin 22:1-13.

Greathead, D. J. 1986. V. Parasitoids used in classical biological control. pp. 304-318 In J Waage D Greathead (Eds), Insect Parasitoids, Academic Press, New York.

Gressitt, J. L., K. P. Rennell, and K.A.J. Wise. 1964. Insects of Campbell Island ecology. Pacific Insects Monographs 7:515-530.

Griffiths-Jones, S. 2005. RALEE - RNA ALignment Editor in Emacs. Bioinformatics 21:257-259.

Griffiths-Jones S., S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, and A. Bateman 2005. Rfam: Annotating non-coding RNAs in complete genomes. Nucl. Acids Res. 33:D121–D141.

Gusfield, D. 1997. Algorithms on strings, trees, and sequences. Cambridge University Press, Cambridge, Massachusetts.

Gutell R. R., J. C. Lee, and J. J. Cannone. 2002. The accuracy of ribosomal RNA comparative structure models. Curr. Opin. Struct. Biol. 12:301–310.

Hagedorn, G. 2003. DeltaAccess – a SQL interface to DELTA, the description language for taxonomy, implemented in Microsoft Access 97, 2000, and 2002. Version 1.81. http://www.diversityworkbench.net/OldModels/Descriptions/index.html.

Harris, H. L., and J. W. Summerlin. 1984. Parasites of horn fly pupae in east central Texas. Southwest. Entomol. 9:169-173.

Hellén, W. 1964. Die Ismarinen und Belytinen Finnlands (Hymenoptera: Proctotrupoidea). Fauna Fennica 18:1-68.

Hellqvist, S. 1994. Biology of *Synacra* sp (Hym, Diapriidae), a parasitoid of *Bradysia paupera* (Dipt, Sciaridae) in Swedish greenhouses. Zeitschrift für Angewandte Entomologie 117:491-497.

Hennig, W. 1966. Phylogenetic systematics. D. D. Davis and R. Zangerl translators (1999). University of Illinois Press, Chicago.

Hickson R. E., C. Simon, and S. W. Perrey. 2000. The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. Mol. Biol. Evol. 17:530-539.

Hines, H. H., J. H. Hunt, T. K. O'Connor, J. J. Gillespie, and S.A. Cameron. 2007. Multigene phylogeny reveals eusociality evolved twice in vespid wasps. P.N.A.S. 104:3295–3299.

Hoffmeister, T. 1989. Biologie und wirtskreis parasitischer Hautflügler der familie Diapriidae. Natur und Museum 119:327-334.

Hogsette, J. A., R. S. Patterson, and P.B. Morgan. 1990. Life history of *Trichopria stomoxydis* (Hymenoptera: Proctotrupoides: Diapriidae) a gregarious endoparasite of *Stomoxys calcitrans* from Zimbabwe. Fla. Entomol. 73:496-502.

Hogsette, J. A., R. Farkas, and R. R. Coler. 1994. Hymenopteran pupal parasites recovered from house fly and stable fly (Diptera: Muscidae) pupae collected on livestock and poultry facilities in northern and central Hungary. Environ. Entomol. 23:778-781.

Hogsette, J. A., R. Farkas, and C. Thuróczy. 2001. Hymenopteran pupal parasitoids recovered from house fly and stable fly (Diptera:Muscidae) pupae collected on livestock Facilities in southern and eastern Hungary. Biol. Control 30:107-111.

Holmes, I. 2004. A probabilistic model for the evolution of RNA structure. BMC Bioinformatics 5:166 (W1-W23).

Holmes I. 2005. Accelerated probabilistic inference of RNA structure evolution. BMC Bioinformatics 6:73.

Honda, M. 1968. Study on the Japanese *Aneurhynchus* Westwood (Hymenoptera: Diapriidae). Journal of the Faculty of Agriculture, Kyushu University 14:613-627.

Honda, M. 1969. Descriptions of two new species of Diapriidae (Hymenoptera) parasitic on *Eristalis* sp and *Lathyrophthalmus ocularis* Coquillett (Diptera: Syrphidae). Mushi 42:155-162.

Hudelot C. V., V. Gowri-shankar, H. Jow., M. Rattray., and P. G. Higgs. 2003. RNA-based phylogenetic methods: Application to mammalian mitochondrial RNA sequences. Mol. Phylogenet. Evol. 28:241-252.

Hudson, G. V. 1892. Note on *Tanyzonous bolitophile*, Marshall, parasitic in the New Zealand Glow-Worm. Entomol. Mon. Mag. 28:277-278.

Huggert, L. 1979. *Cryptoserphus* and Belytinae wasps (Hymenoptera, Proctotrupoidea) parasitizing fungus- and soil-inhabiting Diptera. Notulae Entomologicae 59:139-144.

Huggert, L., and L. Masner. 1983. A review of myrmecophilic-symphilic diapriid wasps in the Holarctic realm, with descriptions of new taxa and a key to genera (Hymenoptera: Proctotrupoidea: Diapriidae). Contributions of the American Entomological Institute 20:63-89.

Huggert, L., and P. B. Morgan. 1993. Description and biology of *Trichopria painteri* n sp (Hymenoptera: Diapriidae), a solitary parasitoid of *Stomoxys calcitrans* (Diptera: Muscidae) from Harare, Zimbabwe. Medical and Veterinary Entomology 7:358-362.

Imms, A. D. 1930. Observations on some parasites of *Oscinella frit* Linnaeus. Parasitology 22:11-36.

International Commission on Zoological Nomenclature. 1999. International Code of Zoological Nomenclature, 4th edition. The International Trust for Zoological Nomenclature, London.

Jacob, S. A. 1980. Pests of ginger and turmeric and their control. Pesticides 14:36-40.

Jaynes, H. A. 1933. The parasites of the Sugarcane Borer in Argentina and Peru, and their introduction into the United States. United States Department of Agriculture Technical Bulletin 363:1-26.

Jeon, Y.-S., H. Chung, S. Park, I. Hur, J.-H. Lee, and J. Chun. 2005. jPHYDIT: a JAVA-based integrated environment for molecular phylogeny of ribosomal RNA sequences. Bioinformatics 21:3171-3173.

Jervis, M. A. 1979. Parasitism of *Aphelopus* species (Hymenoptera: Dryinidae) by *Ismarus dorsiger* (Curtis) (Hymenoptera: Diapriidae). Entomol. Gaz. 30:127-129.

Jiron, L. F., and R. G. Mexzon. 1989. Parasitoid hymenopterans of Costa Rica: Geographical distribution of the species associated with fruit flies (Diptera: Tephritidae). Entomophaga 34:53-60.

Johnston, T. H., and O. W. Tiegs. 1922. On the biology and economic significance of the chalcid parasites of Australian sheep maggot-flies. Proceedings of the Royal Society of Queensland 33:99-128.

Johnson, N. 1992. Catalog of world Proctotrupoidea excluding Platygastridae. Memoirs of the American Entomological Institute, 51:1-825.

Johnson, N. 2003. Hymenoptera On-Line Database. http://iris.biosci.ohio-state.edu/hymenoptera/hym_db_form.html.

Jones, C. M., and D. W. Anthony. 1964. The Tabanidae (Diptera) of Florida. United States Department of Agriculture Technical Bulletin 1295:54-59.

Jow H., C. Hudelot, M. Rattay, and P.G. Higgs. 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Mol. Biol. Evol. 19:1591-1601.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059-3066.

Katoh, K., K. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511-518.

Kauff, F., J. Miadlikowska, and F. Lutzoni. 2003. ARC - a program for ambiguous regions coding. Dept. of Biology, Duke University, USA. http://www.lutzonilab.net/pages/download.shtml.

Kawanishi, M., and T. K. Watanabe. 1981. A parasitic wasp of *Drosophila*. National Institute of Genetics Japan Annual Report 31:107-108.

Kazimirova, M., and V. Vallo. 1992. Influence of larval density of Mediterranean fruit fly (*Ceratitis capitata*, Diptera, Tephritidae) on parasitization by a pupal parasitoid, *Coptera occidentalis* (Hymenoptera, Proctotrupoidea, Diapriidae). Acta Entomologica Bohemoslovaca, 89:179-185.

Kazimirova, M., M. Slovak, and A. Manova. 1997. Host-parasitoid relationship of *Ceratitis capitata* (Diptera: Tephritidae) and *Coptera occidentalis* (Hymenoptera: Proctotrupoidea: Diapriidae) under host heavy metal stress. Eur. J. of Entomol. 94:409-420.

Kazimirova, M., and V. Vallo. 1999. Larval morphology and development of *Coptera occidentalis*. BioControl (Dordrecht), 44:263-280.

Kazimirova, M., and J. Ortel. 2000. Metal accumulation by *Ceratitis capitata* (Diptera) and transfer to the parasitic wasp *Coptera occidentalis* (Hymenoptera). Environmental Toxicology and Chemistry 19:1822-1829.

Kennedy, J. B., R. Kukla, and T. Paterson. 2005. Scientific names are ambiguous as identifiers for biological taxa: Their context and definition are required for accurate data integration. Pages 80-96 in Data integration in the life sciences. Proceedings of the Second International Workshop, San Diego, CA, USA (B. Ludäscher, and L. Raschid, eds.). Springer-Verlag, Berlin.

Kennedy, J., R. Hyam, R. Kukla, and T. Paterson. 2006. Standard data model representation for taxonomic information. OMICS A Journal of Integrative Biology 10:2.

Kieffer, J. J. 1910. Fam. Belytidae. Genera Insectorum, 107:1-47.

Kieffer, J. J. 1916. Diapriidae. Das Tierreich. Vol. 44.. Walter de Gruyter & Co., Berlin, 627 pp.

Kjer, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. 4:314–330.

Kjer, K. M. 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53:506-514.

Kjer K. M., J. J. Gillespie and K. A. Ober. 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. Syst. Biol. 56:133-146.

Kieffer, J. J. 1911. Description de trois nouveaux Diapriides (Hym). Bulletin de la Société Entomologique de France 1911:385-386.

Knutson, L. V., and C. O. Berg. 1963. *Phaenopria popei* (Hymenoptera: Diapriidae) reared from puparia of sciomyzid flies. Can. Ent. 95:724-726.

Lachaud, J. P., and L. Passera. 1982. Donnees sur la biologie de trois Diapriidae myrmecophiles: *Plagiopria passerai* Masner *Solenopsia imitatrix* Wasmann et *Lepidopria pedestris* Kieffer. Insectes Sociaux 29:561-568.

Lambkin, C. L., M. S. Y. Lee, S. L. Winterton, and D. K. Yeates. 2002. Partitioned Bremer support and multiple trees. Cladistics 18:436-444.

Larissa, G., M. Aluja, E. Miguel, and J. Sivinski. 2002. Performance of two fruit fly (Diptera: Tephritidae) pupal parasitoids (*Coptera haywardi* (Hymenoptera: Diapriidae) and *Pachycrepoideus vindemiae* (Hymenoptera: Pteromalidae)) under different environmental soil conditions. Biol. Control 23:219-227.

Larsen, N., G. J. Olsen, B. L. Maidak, M. J. McCaughey, R. Overbeek, T. J. Macke, T. L. Marsh, and C. R. Woese. 1993. The ribosomal database project. Nucleic Acids Res. 21:3021-3023.

Lee, M. S. Y. 2001. Unalignable sequences and molecular evolution. Trends Ecol. Evol. 16:681–685.

Legner, E. F. 1966. Parasites of the House Fly and other filth-breeding Diptera in Southern California. J. Econ. Entomol. 59:999-1001.

Legner, E. F., E. C. Bay, and E. B. White. 1967. Activity of parasites from Diptera*: Musca domestica*, *Stomoxy calcitrans*, *Fannia canicularis* and *F. femoralis*, at sites in the Western Hemisphere. Ann. Entomol. Soc. Am. 60:462-468.

Legner, E. F., and G. S. Olton. 1968. Activity of parasites from Diptera: *Musca domestica*, *Stomoxys calcitrans*, and species of *Fannia*, *Muscina* and *Ophyra* II, at sites in the Eastern Hemisphere and Pacific area. Ann. Entomol. Soc. Am. 61:1306-1314.

Legner, E. F., and D. J. Greathead. 1969. Parasitism of pupae in East African populations of *Musca domestican* and *Stomoxys calcitrans*. Ann. Entomol. Soc. Am. 62:128-133.

Legner, E. F., and G. S. Olton. 1971. Distribution and relative abundance of dipterous pupae and their parasitoids in accumulations of domestic animal manure in the southwestern United States. Hilgardia 40:505-535.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913-925.

Lindgreen, S., P. P. Gardner, and A. Krogh. 2006. Measuring covariation in RNA alignments: Physical realism improves information measures. Bioinformatics 22:2988-2995.

Loiácono, M. S. 1981a. Diapridos ecitofilos sinfilos de la region Neotropical (Hymenoptera - Proctotrupoidea - Diapriidae). Revista de la Sociedad Entomologica Argentina 40:297–310.

Loiácono, M. S. 1981b. Notas sobre Diapriinae neotropicales (Hymenoptera, Diapriidae). Revista de la Sociedad Entomológica Argentina 40:237-241.

Loiácono, M. S. 1987. Un nuevo diaprido (Hymenoptera) parasitoide de larvas de *Acromyrmex ambiguus* (Emery) (Hymenoptera, Formicidae) en el Uruguay. Revista de la Sociedad Entomologica Argentina 44:129–136.

Loiácono, M. S. 2000. Diaprinos asociados a la hormiga *Camponotus rufipes* (Hymenoptera: Diapriidae). Revista de la Sociedad Entomologica Argentina 59:198–200.

Loiácono, M. S. and C. B. Margaria. 2000. Systematic and cladistic analysis of the myrmecophilic genus *Gymnopria* Loiácono (Hymenoptera: Diapriidae). Insect Systematics and Evolution 31:187-200.

Loiácono, M. S., C. B. Margaria, E. M. Quiran, and B. M. Corró-Molas. 2000. Diapridos (Hymenoptera) parasitoides de larvas de la hormiga cortadora *Acromyrmex lobicornis* (Hymenoptera: Formicidae) en la Argentina. Revista de la Sociedad Entomológica Argentina 59:7-15.

Loiácono, M. S., and C. B. Margaria. 2002. Ceraphronoidea, Platygastroidea, and Proctotrupoidea from Brazil (Hymenoptera). Neotropical Entomology 31:551-560.

López, M., M. Aluja, and J. Sivinski. 1999. Hymenopterous larval-pupal and pupal parasitoids of *Anastrepha* flies (Diptera: Tephritidae) in Mexico. Biol. Control 15:119-129.

Lopez, M. O. 1996. Patrones de parasitismo en moscas de la fruta del genero *Anastrepha* (Diptera: Tephritidae), en frutos nativos (*Spondias mombin* L y *Psidium guajava* L) y exotocos (*Magifera indica* L y *Citrus sinensis* L). Univ. of Veracruz, Xalapa.

Löytynoja, A. and M. C. Milinkovitch. 2001. SOAP, cleaning multiple alignments from unstable blocks. Bioinformatics 17:573-574.

Lutzoni F., P. Wagner, V. Reeb, and S. Zoller. 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49:628-651.

Macek, J. 1989a. Studies on the Diapriidae (Hymenoptera, Proctotrupoidea). Part 1. Taxonomic remarks on the subfamily Belytinae with particular reference to the tribus Pantolytini. Annales Zoologici, Polska Akademia Nauk Instytut Zoologii, 42:353-362.

Macek, J. 1989b. Studies on the Diapriidae (Hymenoptera, Proctotrupoidea). Part 2. A revision of the genus *Anommatium* Foerster. Annales Zoologici, Polska Akademia Nauk Instytut Zoologii, 43:127-132.

Macek, J. 1993a. Revision of Holarctic *Polypeza* (Hymenoptera, Diapriidae). Folia Heyrovskyana, 1:19-24.

Macek, J. 1993b. Revision of European *Pantolyta* Foerster (Hymenoptera, Diapriidae). Folia Heyrovskyana, 1:41-51.

Macek, J. 1994. The origin of ventral bridge in Diapriidae (Hymenoptera,

Proctotrupoidea) with comments on morphology of head posterior. Folia Heyrovskyana, 2:21-26.

Macek, J. 1995a. Revision of west Palaerctic *Lyteba* Th. (= *Oxylabis* auct.) (Hymenoptera: Proctotrupoidea, Diapriidae). Folia Heyrovskyana, 3:29-39.

Macek, J. 1995b. A taxonomic revision of European *Psilommina* (Hymenoptera: Diapriidae) Part 2. The *Synacra* complex. European Journal of Entomology, 92:469-482.

Macek, J. 1995c. Revision of genus *Opazon* Haliday in Europe (Hymenoptera, Diapriidae). Folia Heyrovskyana, 3:80-87.

Macek, J. 1995d. Revision of the European species of *Belyta* Jurine. Acta musei Nationalis Pragae, Series B, Historia Naturalis, 51:1-22.

Macek, J. 1996. Description of new species of *Synbelyta* Hellen (Hymenoptera, Diapriidae) with taxonomical remarks on the genus. Casopis Národního Muzea Rada Prírodovedná, 165:63-74.

Macek, J. 1997a. Description of new species of *Aprestes* (Diapriidae, Hymenoptera) with taxonomic remarks on the genus. Casopis Národního Muzea Rada Prírodovedná, 166:39-45.

Macek, J. 1997b. Revision of central European species of *Zygota* (Hymenoptera: Diapriidae) with description of new species and taxonomic remarks on the genus. Folia Heyrovskyana, 5:31-57.

Macek, J. 1997c. Revision of the genus *Macrohynnis* (Hymenoptera: Diapriidae) with description of new species. Folia Heyrovskyana, 5:139-146.

Macek, J. 1998. Revision of European species of *Acropiesta* (Hymenoptera: Diapriidae). Folia Heyrovskyana, 6:21-42.

Mackenzie, C. L., and J. V. Richerson. 1993. Parasitoids of the horn fly in rangeland ecosystems of trans-pecos Texas. Southwest. Entomol. 18:57-59.

Maddison, D. R., and K.-S. Schulz (eds.) 1996-2007. The Tree of Life web project. http://tolweb.org.

Magnarelli, L. A., and J. F. Anderson. 1980. Parasitism of *Tabanus nigrovittatus* immatures (Diptera: Tabanidae) by *Trichopria* sp (Hymenoptera: Diapriidae). J. of Medical Entomology 17:481-482.

Maier, C. T. 1981. Parasitoids emerging from puparia of *Rhagoletis pomonella* (Diptera: Tephritidae) infesting hawthorn and apple in Connecticut. Can. Ent. 113:867-870.

Mann, W. M., (1914) Some myrmecophilic insects from Mexico. Psyche, 21, 171–184.

Marchiori, C. H., A. T. Oliviera, and A. X. Linhares. 2000. *Trichopria* sp (Hymenoptera: Diapriidae) parasitoides de Diptera Muscoidea. Archivos do Instituto Biológico Sao Paulo 67:131-133.

Marchiori, C. H. 2001. Primeira ocorrencia do parasitoide *Trichopria* sp (Hymenoptera: Diapriidae) em pupas de *Haematobia irritans* (L) (Diptera: Muscidae) no Brasil. Entomologia y Vectores 8:281-285.

Marshall, T. A. 1892. On a new genus and species of Belytidae from New Zealand. Entomologist's Monthly Magazine 28:275-276.

Masner, L. 1961. Ambositrinae, a new subfamily of Diapriidae from Madagascar and central Africa (Hymenoptera Proctotrupoidea). Memoires de l'Institut Scientifique de Madagascar, 12:289-295.

Masner, L. 1969. The geographic distribution of recent and fossil Ambositrinae (Hymenoptera, Proctotrupoidea: Diapriidae). Tagungsberichte, 80:105-109.

Masner, L. 1976. A revision of the Ismarinae of the New World (Hymenoptera, Proctotrupoidea, Diapriidae). Canadian Entomologist, 108:1243-1266.

Masner, L. 1991. Revision of *Spilomicrus* Westwood in America north of Mexico (Hymenoptera: Proctotrupoidea, Diapriidae). Can. Ent. 123:107-177.

Masner, L. 1993. Superfamily Proctotrupoidea. pp. 537-557 In J.T. Huber, H. Goulet (Eds), Hymenoptera of the world: An identification guide to families, Agriculture Canada, Ottawa, 668 pp.

Masner, L., and J. L. García. 2002. The genera of Diapriinae (Hymenoptera: Diapriidae) in the New World. Bulletin of the American Museum of Natural History, 268:1-138.

Mathews, D. H., and D. H. Turner. 2002. Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. J. Mol. Biol. 317:191-203.

McPhail, M., and C. I. Bliss. 1933. Observations on the Mexican Fruit Fly and some related species in Cuernavaca, Mexico, in 1928 and 1929. United States Department of Agriculture Circular 255:1-24.

Megyaszai, T., and C. Thuróczy. 1998. The papers (1956-1986) published by J. B. Szabó on Hymenoptera and his described taxa. Rovartani Közlemények [=Folia Entomologica Hungarica], 59:185-208.

Mendes, J., and A. X. Linhares. 1999. Diapause, pupation sites and parasitism of the horn fly, *Haematobia irritans*, in south-eastern Brazil. Medical and Veterinary Entomology 13:185-190.

Miadlikowska, J., F. Lutzoni, T. Goward, D. Zoller, and D. Posada. 2003. New approach to an old problem: Gap-rich regions from ITS and rDNA large-subunit are incorporated into phylogenetic analyses to resolve the *Peltigera canina* species complex. Mycologia 95:1181-1203.

Mitchell, E. R., W. W. Copeland, and F. C. Tingle. 1974. Parasites of filth-breeding diptera in poultry houses in north central Florida. Fla. Entomol. 57:383-384.

Morgenstern, B., S. J. Prohaska, D. Pöhler, and P. F. Stadler. 2006. Multiple sequence alignment with user-defined anchor points. Algorithms for Molecular Biology 1:6.

Morris, R. A., M. Passell, Wan, J., R. D. Stevenson, and W. Haber. 2002. Engineering considerations for biodiversity software. Pages 49-59 *in* Towards a global biological information infrastructure Challenges, opportunities, synergies, and the role of entomology (H. Saarenmaa and E. S. Nielsen eds.). Technical Report. European Environment Agency.

Morrison, D. A. 2006. Multiple sequence alignment for phylogenetic purposes. Australian Systematic Botany 19:479–539.

Morrison, D. A., and J. T. Ellis. 1997. Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of Apicomplexa. Mol. Biol. Evol. 14:428-441.

Muesebeck, C. F. W. 1949. A new flightless *Phaenopria*. Can. Ent. 81:234-235.

Muesebeck, C. F. W. 1958. Superfamily Proctotrupoidea. Pages 88-94 *in* Hymenoptera of America North of Mexico Synoptic Catalog (Agriculture Monograph No. 2), First Supplement (K. V. Krombein ed.). United States Government Printing Office, Washington, DC.

Muesebeck, C. F. W. 1961. A new Japanese *Trichopria* parasitic on the house fly (Hymenoptera:Diapriidae). Mushi 35:1-2.

Muesebeck, C. F. W. 1980. The Nearctic parasitic wasps of the genera *Psilus* Panzer and *Coptera* Say (Hymenoptera, Proctotrupoidea, Diapriidae). United States Department of Agriculture Technical Bulletin(1617) pp 1-71.

Muesebeck, C. F. W., and L. M. Walkley. 1951. Superfamily Proctotrupoidea. Pages 655-718 *in* Hymenoptera of America north of Mexico -- Synoptic Catalog. (C. F. W. Muesebeck, K. V. Krombein, and H. K. Townes eds), U.S. Dept. Agriculture Monograph No. 2. 1420 pp.

Müller, J., and R. R. Reisz. 2006. The phylogeny of early Eureptiles: Comparing parsimony and Bayesian approaches in the investigation of a basal fossil clade. Systematic Biology, 55:3 503-511.

Munk, T. 1991. Two species of Diapriinae (Diapriidae, Hymenoptera) reared from *Tephrochlamys tarsalis* (Heleomyzidae, Diptera). Flora og Fauna 97:47-49.

Musetti, L. F., and N. F. Johnson. 2004. Revision of the New World species of the genus *Monomachus* Klug (Hymenoptera: Proctotrupoidea, Monomachidae). Canadian Entomologist, 136:501-552.

Myers J. G. 1931. A preliminary report on an investigation into the biological control of West Indian insect pests. H. M. Stationery Office, London. 178 pp.

Myers, J. G. 1935. Second report on an investigation into the biologiocal control of West Indian insect pests. Bulletin of Entomological Research 26:181-252.

Narayanan, E. S., and S. S. Chawla. 1962. Parasites of fruit fly pests of the World. Beitraege zur Entomologie 12:347-476.

Naumann, I. D. 1982. Systematics of the Australian Ambositrinae (Hymenoptera: Diapriidae), with a synopsis of the non-Australian genera of the subfamily. Australian Journal of Zoology, Supplementary Series No. 85:1-239.

Naumann, I. D. 1987. The Ambositrinae (Hymenoptera: Diapriidae) of Melanesia. Invertebrate Taxonomy, 1:439-471.

Naumann, I. D. 1988. Ambositrinae (Insecta: Hymenoptera: Diapriidae). Fauna of New Zealand, Supplementary Series No. 15:1-165.

Naumann, I. D. and L. Masner. 1980. A revision of the termitophilous Australian genus *Leaiopria* Dodd (Hymenoptera: Proctotrupoidea: Diapriidae). Journal of the Australian Entomological Society, 19:143-149.

Naumann, I. D. and L. Masner. 1985. Parasitic wasps of the proctotrupoid complex: a new family from Australia and a key to world families (Hymenoptera: Proctotrupoidea sensu lato). Australian Journal of Zoology, 33:761-783.

Nixon, G. E. J. 1930. The Ethiopian representatives of the genus *Galesus* (Proctotrupoidea), with descriptions of new species. Ann. Mag. Nat. Hist. 10:399-414.

Nixon, G. E. J. 1940. A new African diapriid (Hym, Proctotrupoidea). Bull. Entomol. Res. 31:59-60.

Nixon, G. E. J. 1947. A new myrmecophilous diapriid (Hymenoptera Serphoidea). Annals and Magazine of Natural History, 13, 787–789.

Nixon, G. E. J. 1957. Hymenoptera, Proctotrupoidea, Diapriidae, subfamily Belytinae. Handbooks for the Identification of British Insects 8(3dii):1-107.

Nixon, G. E. J. 1980. Diapriidae (Diapriinae). Hymenoptera, Proctotrupoidea. Handbooks for the identification of British Insects, 8:1-55.

Nixon, K. C., J. M. Carpenter, and S. J. Borgardt. 2001. Beyond NEXUS: Universal cladistic data objects. Cladistics 17:S53-S59.

Noller, H. F., and C. R. Woese. 1981. Secondary structure of 16S ribosomal RNA. Science 212:403-411.

Notredame, C., E. A. O'Brien, and D. G. Higgins. 1997. RAGA: RNA sequence alignment by genetic algorithm. Nucleic Acids Research. 25:4570–4580.

Notton, D. G. 1991. Some Diptera host records for species of *Basalys* and *Trichopria* (Hym, Diapriidae). Entomol. Mon. Mag. 127:123-126.

Notton, D. G. 1994. New and notable European *Trichopria* Ashmead, 1893 (Hym, Proctotrupoidea, Diapriidae). Entomologiske Meddelelser 62:9-12.

Notton, D. G. 1997. *Synacra paupera* Macek (Hym, Diapriidae) new to Britain: A parasitoid of the greenhouse pest *Bradysia paupera* Tuomikoski (Dipt, Sciaridae). Entomol. Mon. Mag. 133:257-259.

Notton, D. G. 2004. A catalogue of types of Diapriinae (Hymenoptera, Diapriidae) at the National Museum of Natural History, Paris, with notes on the classification of Diapriinae and a brief history of the types of Jean-Jacques Kieffer (1856-1925). Zoosystema, 26: 315-352.

Offenberger, M., and A. J. Klarenberg. 1994. Fitting parasitoid wasps to their drosophilid hosts: An overview of collecting methods with a checklist of hymenopteran parasitoids and their drosophilid hosts in European temperate woodlands. Drosophila Information Service 75:123-125.

Ogden, T. H., and M. F. Whiting. 2005. Phylogeny of Ephemeroptera (mayflies) based on molecular evidence. Mol. Phylogenet. Evol. 37:625–643.

Ogden, T. H., M. F. Whiting, and W. C. Wheeler. 2005. Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: A reply to Kjer. Cladistics 21:295-302.

O'Leary, M. A., and S. G. Kaufman. 2007. MorphoBank 2.5: Web application for morphological phylogenetics and taxonomy. http://www.morphobank.org.

O'Neill, W. L. 1973. Biology of *Trichopria popei* and *T. atrichomelinae* (Hymenoptera: Diapriidae), parasitoids of Sciomyzidae (Diptera). Ann. Entomol. Soc. Am. 66:1043-1050.

Osborn, A. W. 1974. Predators, parasitoids and pathogens of the Sugarcane Soldier Fly, *Inopus rubriceps* (Macquart). Proceedings of the International Society of Sugar Cane Technologists 15:423-427.

Osborn, A. W., G. N. R. Forteath, and G. A. Holloway. 1973. A new species of *Neurogalesus* (Hymenoptera: Diapriidae), a parasite of the soldier fly, *Inopus rubriceps* (Diptera: Stratiomyidae). J. Aust. Entomol. Soc. 12:124-126.

Ovruski, S., M. Aluja, J. Sivinski, and R. Wharton. 2000. Hymenopteran parasitoids on fruit-infesting Tephritidae (Diptera) in Latin America and the southern United States: Diversity, distribution, taxonomic status and their use in fruit fly biological control. Integrated Pest Management Reviews 5:81-107.

Page, R. D. M. 2000. Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. Nucleic Acids Research 28:3839-3845.

Page, R. D. M. 2005. A taxonomic search engine: Federating taxonomic databases using web services. BMC Bioinformatics 6:48.

Paulson, G. S., and R. D. Akre. 1991. *Trichopria* sp (Hymenoptera: Diapriidae) reared from *Microdon albicomatus* Novak (Diptera: Syrphidae). Can. Ent. 123:719.

Pemberton, C. E., and H. F. Willard. 1918. A contribution to the biology of fruit-fly parasites in Hawaii. Journal of Agricultural Research 15, 419-465.

Philippe, H., F. Delsuc, H. Brinkmann, and N. Lartillot. 2005. Phylogenomics. Annual Review of Ecology, Evolution, and Systematics 36:541-562.

Phillips, A. J. 2006. Homology assessment and molecular sequence alignment. Journal of Biomedical Informatics 39:18–33.

Polaszek, A. 2005. A universal register for animal names. Nature 437:477.

Pride, D. T., R. J. Meinersmann, T. M. Wassenaar, and M. J. Blaser. 2003. Evolutionary implications of microbial genome tetranucleotide frequency biases. Genome Res. 13:145-158.

Pullan, M. R., K. E. Armstrong, T. Paterson, A. Cannon, J. B. Kennedy, M. F. Watson, S. McDonald, and C. Raguenaud. 2005. The Prometheus Description Model: An examination of the taxonomic description-building process and its representation. Taxon 54:751-765.

Puttarudriah, M., and S. Usman. 1954. Mysore Dacinae and their natural parasites. Mysore Agricultural Journal 30:257-260.

Qi, J., H. Luo, and B. Hao. 2004a. CVTree: A phylogenetic tree reconstruction tool based on whole genomes. Nucleic Acids Res. 32:W45–W47.

Qi, J., B. Wang, and B. I. Hao. 2004b. Whole proteome prokaryote phylogeny without sequence alignment: A k-string composition approach. Journal of Molecular Evolution 58:1-11.

Raga, A., D. A. O. Prestes, M. F. Souza Filho, M. E. Sato, R. C. Siloto, J. A. Guimaraes, and R. A. Zucchi. 2004. Fruit fly (Diptera: Tephritoidea) infestation in citrus in the state of São Paulo, Brazil. Neotropica 33:85-89.

Remsen, D. P., C. Norton, and D. J. Patterson. 2006. Taxonomic informatics tools for the electronic Nomenclator Zoologicus. Biol Bull. 210:18-24.

Robertson, L. N. 1987. The effects of natural enemies on the population dynamics of sugarcane soldier fly, *Inopus rubriceps* (Dip: Stratiomyidae). Agriculture Ecosystems and Environment 19:343-363.

Robertson, L. N., and M. P. Zalucki. 1985. The natural enemies of Australian (sugarcane) soldier fly, *Inopus rubriceps* (Diptera: Stratiomyidae). Proceedings of the 4th Australasian Conference on Grassland Invertebrate Ecology pp. 228-235.

Rodman, J. E., and Cody J. H. 2003. The taxonomic impediment overcome: NSF's partnerships for enhancing expertise in taxonomy (PEET) as a model. Systematic Biology 52:428-435.

Romani, R., N. Isidoro, F. Bin, and B. Vinson. 2002. Host recognition in the pupal parasitoid *Trichopria drosophilae*: A morpho-functional approach. Entomologia Experimentalis et Applicata 105:119-128.

Ronquist, F., A. P. Rasnitysn, A. Roy, K. Eriksoson, and M. Lindgren. 1999. Phylogeny of the Hymenoptera: A cladistic reanalysis of Rasnitsyn's (1988) data. Zoologica Scripta, 28:13-50.

Ronquist, F., and J. P. Huelsenbeck, J.P. 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Sanders, G. E. 1911. Notes on the breeding of *Tropidopria conica* Fabr. Can. Ent. 43:48-50.

Schluenzen, F., A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, and F. Franceschi. 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. Cell 102:615-623.

Segal, B. 1936. Synopsis of the Tabanidae of New York, their biology and taxonomy. J. N. Y. Entomol. Soc. 44:51-78.

Seibel. P. N., T. Müller, T. Dandekar, S. Schultz, and M. Wolf. 2006. 4SALE – A tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinformatics. 7:498.

Sharkey, M. J., and A. Roy. 2002. Phylogeny of the Hymenoptera: A reanalysis of the Ronquist et al. (1999) reanalysis, emphasizing wing venation and apocritan relationships. Zoologica Scripta, 31:57-66.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114-1116.

Silvestri, F. 1913. Viaggio in Africa per cercare parassiti di mosche dei frutti. Bollettino del Laboratorio de Zoologia Generale e Agraria dell Facoltà Agraria in Portici, 8:1-164.

Silvestri, F. 1914. Report of an expedition to Africa in search of the natural enemies of fruit flies (Trypaneidae), with descriptions, observations and biological notes. Bull. Ent. Agric. Hawaii 3:1-176.

Simmonds, F. J. 1953. Parasites of the fruit-fly *Oscinella frit* (L.) in Eastern North America. Bull. Entomol. Res. 43:503-542.

Sinclair, B. J. 2000. Imature stages of Australian *Austrothaumalea* Tonnoir and *Niphta* Theischinger (Diptera: Thaumauleidae). Aust. J. Entomol. 39:171-176.

Sivinski, J., K. Vulinec, E. Menezes, and M. Aluja. 1998. The bionomics of *Coptera Haywardi* (Ogloblin) (Hymenoptera: Diapriidae) and other pupal parasitoids of tephritid fruit flies (Diptera). Biol. Control 11:193-202.

Sivinski, J., J. Piñero, and M. Aluja. 2000. The distributions of parasitoids (Hymenoptera) of *Anastrepha* Fruit Flies (Diptera: Tephritidae) along an altitudinal gradient in Veracruz, Mexico. Biol. Control 18:258-269.

Skovgard, H., and J. B. Jespersen. 1999. Activity and relative abundance of hymenopterous parasitoids that attack puparia of *Musca domestica* and *Stomoxys calcitrans* (Diptera:Muscidae) on confined pig and cattle farms in Denmark. Bulletin of Entomological Research 89:263-269.

Smith, A. D., T. W. H. Lui, and E. R. M. Tillier. 2004. Empirical models for substitution in ribosomal RNA. Mol. Biol. Evol. 21:419-427.

Smith, J. P., R. D. Hall, and G. D. Thomas. 1987. Field parasitism of the stable fly (Diptera: Muscidae). Ann. Entomol. Soc. Am. 80:391-397.

Smythe, A. B., M. J. Sanderson, and S.A. Nadler. 2006. Nematode small subunit phylogeny correlates with alignment parameters. Syst. Biol. 55:972-992.

Sneath, P. H. A., and R. R. Sokal. 1973. Numerical Taxonomy: The principles and practice of Numerical Classfication. W. H. Freeman, San Fransisco.

Soenarjo, E. 1986. Parasitoids of the Rice Gall Midge (GM) in Indonesia. International Rice Research Newsletter 11:29.

Spahn, C. M., R. Beckmann, N. Eswar, P. A. Penczek, A. Sali, G. Blobel, and J. Frank. 2001. Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions. Cell 107:373-386.

Stibick, J. N. L. 2004. Natural enemies of true fruit flies (Tephritidae). United States Department of Agriculture Bulletin, pp. 1-86.

Stuart, G. W., and M. W. Berry. 2004. An SVD-based comparison of nine whole eukaryotic genomes supports a coelomate rather than ecdysozoan lineage. BMC Bioinformatics 5:204.

Szelényi, G. 1957. Ujabb adatok az amerikai feher szovolepte eloskodoinek ismeretehez [Some new data on the hymenopterous parasites of Hyphantria cunea Drury]. Annales Instituti Protectionis Plantarum Hungarici 7:295-312.

Taley, Y. M., and K. R. Thakare. 1979. Biology of seven new hymenopterous parasitoids of *Atherigona soccata* (Rondani). Indian Journal of Agricultural Science 49:344-354.

Taylor, R. W. 1983. Descriptive taxonomy: Past, present, and future. Pages 93–134 *in* Australian systematic entomology: A bicentenary perspective (E. Highley, and R. W. Taylor, eds.). CSIRO, Canberra.

Telford, M. J., M. J. Wise, and V. Gowri-Shankar. 2005. Consideration of RNA Secondary Structure Significantly Improves Likelihood-Based Estimates of Phylogeny: Examples from the Bilateria. Mol. Biol. Evol. 22:1129-1136.

Teodorescu, I., and A. Ursu. 1979. [Species of Diapriinae (Hymenoptera - Proctotrupoidea), puparia parasites of synantropic Diptera]. Studii si Cercetari de Biologie Seria Biologie Animala 31:131-136.

Thakur, B., and S. Odak. 1982. New record of the parasites of *Melanagromyza obtusa* Malloch. Science and Culture 48:80.

Thomas, G. D., and C. E. Morgan. 1972. Parasites of the horn fly in Missouri. J. Econ. Entomol. 65:169-174.

Thompson, F. C. (editor). 2005. Biosystematic Database of World Diptera, Version 7.5. http://www.diptera.org/names, accessed on 2005-2007.

Thompson, W. R. 1955. A catalogue of the parasites and predators of insect pests. Section 2. Host parasite catalogue, Part 3. Hosts of the Hymenoptera (Calliceratid to Evaniid). Commonwealth Agricultural Bureaux, The Commonwealth Institute of Biological Control, Ottawa, Ontario, Canada.

Ulitsky, I., D. Burstein, T. Tuller, and B. Chor. 2006. The average common substring approach to phylogenomic reconstruction. Journal of Computational Biology, 13:336–350.

Ulrich, W. 2000. Parasitic Hymenoptera and their hosts in a beech forest on limestone: Dependence on biotic and abiotic factors. Polskie Pismo Entomologiczne, 69:429-439.

Vallo, V. 1989. Control of the European cherry fruit fly *Rhagoletis cerasi* (L.) in Chechoslovakia. Tagungsberichte, 278:281-285.

Van de Peer, Y., J. M. Neeps, P. DeRijk, and R. De Wachter. 1993. Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: Calibration of the molecular clock. J. Mol. Evol. 37, 221–232.

van den Berg, M. A. 1974. Natural enemies and diseases of the forest insect *Nudaurelia cytherea clarki* Geertsema (Lepidoptera:Saturniidae). Phytophylactica 6:39-44.

Vaughan, J. T. 1985. Biology of *Trichopria stomoxydis* (Hymenoptera: Diapriidae) and its potential use as a biological control agent against *Stomoxys calcitrans* (Diptera: Muscidae). Master's Thesis, University of Florida, Gainesville.

Veeranna, G., M. R. Nirmala, and D. Mahadevappa. 1987. Record of a new hyperparasite, *Trichopria* sp (Hymenoptera, Diapriidae) of *Tricholyga bombycis* Beck (Diptera, Tachinidae), a larval parasite of *Bombyx mori* L. Current Science 56:1031-1032.

Vinga, S., and J. Almeida. 2003. Alignment-free sequence comparison-a review. Bioinformatics 19:513-523.

Vukasovic P. 1928. Contribution to the study of entomophagous insect parasites. Glas Srpske Kral. Akad. 131:45-72.

Vukasovic, P. 1926. Biological observations on the hymanopterous parasites of Diptera Brachycera. Glasnik Cent. Hig. Zavoda 1:195-211.

Wall, I. 1967. Die Ismarinae und Belytinae der Schweiz. Entomol. Abh. Staatliches Mus. Tierkunde, Dresden, 35:123-265.

Watts, K. J., and J. R. L. Combs. 1977. Parasites of *Haematopbia irritans* and other flies breeding in bovine feces in northeast Mississippi. Environ. Entomol. 6:823-826.

Waugh, A., P. Gendron, R. Altman, J. W. Brown, D. Case, D. Gautheret, S. C. Harvey, N. Leontis, J. Westbrook, E. Westhof, M. Zuker, and F. Major. 2002. RNAML: A standard syntax for exchanging RNA information. RNA 8:707–717.

Wharton, R. A., M. K. Trostle, R. H. Messing, R. S. Copeland, S. W. Kimani-Njogu, , S. Lux, W. A. Overholt, S. Mohamed, and J. Sivinski. 2000. Parasitoids of medfly, *Ceratitis capitata*, and related tephritids in Kenyan coffee: A predominantly koinobiont assemblage. Bulletin of Entomological Research 90:517-526.

Wharton, R. A., M. J. Yoder, Gillespie, J. J., J. C. Patton, and R. L. Honeycutt. 2006. Relationships of *Exodontiella*, a non-alysiine, exodont member of the family Braconidae (Insecta, Hymenoptera). Zoologica Scripta 35:323-340.

Wheeler, W. C. 1999. Fixed character states and the optimization of molecular sequence data. Cladisitics 15:379-385.

Wheeler W. C., and R. L. Honeycutt. 1988. Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. Mol. Biol. Evol. 5:90-96.

Wheeler, W. C., J. Gatesy, and R. DeSalle. 1995. Elision: Accommodating multiple molecular sequence alignments ambiguous sites. Mol. Phylogenet. Evol. 4:1–9.

Wheeler, W. C., D. Gladstein, and J. De Laet. 1996–2003. POY, Version 3.0. (Current Version 3.0.11). American Museum of Natural History, New York, NY.

Wheeler, W. C., L. Aagesen, C. P. Arango, J. Faivovich, T. Grant, C. D'Haese, D. Janies, W. L. Smith, A. Varón, and G. Giribet. 2006. Dynamic homology and phylogenetic systematics: A unified approach using POY. American Museum of Natural History, New York. 365 pp.

Wimberly, B. T., D. E. Brodersen, W. M. Clemons Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan. 2000. Structure of the 30S ribosomal subunit. Nature 407:327-339.

Whitfield, J. B. 1992. Phylogeny of the non-aculeate Apocrita and the evolution of parasitism in the Hymenoptera. Journal of Hymenoptera Research 1:3-14.

Woese, C. R., R. Gutell, R. Gupta, and H. F. Noller. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiology Reviews 47:621-669.

Wright, D. W., Q. A. Geering, and D. G. Ashby. 1947. The insect parasites of the Carrot Fly *Psila rosae* Fab. Bull. Entomol. Res. 37:507-529.

Wu X., and Lin, G. 2005. Selected string representation for whole genomes. Computational Intelligence in Bioinformatics and Computational Biology. CIBCB '05. Proceedings of the 2005 IEEE Symposium 1-8.

Wu, T-J., Y.-H. Huang, and L.-A. Li. 2005. Optimal word sizes for dissimilarity measures and estimation of the degree of dissimilarity between DNA sequences. Bioinformatics 21:4125-4132.

Wuyts, J., P. De Rijk, Y, Van de Peer, T. Winkelmans, and R. De Wachter. 2001. The European Large Subunit Ribosomal RNA Database. Nucleic Acids Res. 29:175-177.

Wuyts, J., Y. Van de Peer, T. Winkelmans, and R. De Wachter. 2002. The European database on small subunit ribosomal RNA. Nucleic Acids Res. 30:183-185.

Yasumatsu, K. 1964. Diapriidae. Pages 93-94 *in* A tentative catalogue of insect natural enemies of injurious insects in Japan (K. Yasumatsu, and C. Watanabe, C. eds.). Entomological Laboratory Faculty of Agriculture, Kyushu, Fukuoka, Japan.

Yoder, M. J. 2004. Revision of the North American species of the genus *Entomacis* (Hymenoptera: Diapriidae). Can. Ent. 136:323-405.

Yoder, M. 2007. *Mannomicrus* (Hymenoptera: Diapriidae), a new genus of myrmecophilic diapriid, with a digital version of Masner and García's (2002) key to New World Diapriinae and an illustration of digital description and key markup using an ontology. Zootaxa 1439:47-55.

Yoder, M. J., K. Dole, and A. R. Deans. 2006. mx – A collaborative content management system for revisionary systematists. Available following links from http://www.diapriid.org.

Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller. 2001. Crystal structure of the ribosome at 5.5 Å resolution. Science 292:883-896.

Zaldivar-Riverón, A., M. Mori and D. L. J. Quicke. 2006. Systematics of the cyclostome subfamilies of braconid parasitic wasps (Hymenoptera: Ichneumonoidea): A simultaneous molecular and morphological Bayesian approach. Mol. Phylogenet. Evol. 38:130-145.

VITA

Matthew Jon Yoder (Matt) was raised in a small town in north-central rural Saskatchewan, Canada. In 1995 he completed a four year non-major honors science bachelors degree at the University of Waterloo, Waterloo, Ontario. Following some time abroad he found himself at College Station, Texas, where he accepted an assistantship to help develop the Texas A&M University Insect Collection's specimen databases as he studied for his Masters of Science (Texas A&M, Department of Entomology). After completion (2002) of his M.S. on the taxonomy of parasitic wasps of the genus *Entomacis*, Matt began a Ph.D., also at Texas A&M (Department of Entomology), which he completed in 2007. The results of that effort are chronicled here. Matt can be reached at his 'diapriid' gmail email, or at Box 929, Rosthern, Saskatchewan, S0K, 3R0.